

Tue Apr 15 15:15:17 2003

us-09-830-244b-1.rag

Page 1

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 21:24:26 ; Search time 109 Seconds

(Without alignments)
116.136 Million cell updates/sec

Title: US-09-830-244b-1

Perfect score: 522

Sequence: 1 MWFHKVGRGHRKVFWEWD.....NHTNNTNINRINISNCT 95

Scoring table:

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database:

Maximum Match 100%
Listing first 45 summaries

A.Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	100.0	95	21	AAV94526 Human lysine-rich
2	73.5	14.1	234	22	AAU25543 G protein-coupled
3	73	14.0	310	20	AAV37060 Chlamydia trachoma
4	73	14.0	412	23	ABR94350 Chlamydia trachoma
5	73	14.0	426	23	ABR94350 Chlamydia trachoma
6	73	14.0	433	23	ABR94350 Chlamydia trachoma
7	71	13.6	1226	18	AAV13825 Yeast transcript
8	69.5	13.3	1381	23	ABR09493 Ampicillin ABC transpor
9	69	13.2	1245	21	AAV18244 Plasmodium falcipar
10	68	13.0	346	19	AAV98706 H. pylori GHPD 727

11	68	13.0	907	23	ABR49049
12	67	12.8	850	22	ABR65764
13	67	12.8	1028	22	ABR62708
14	66	12.6	244	23	ABR27420
15	65.5	12.5	61	23	ABR07000
16	65	12.5	618	23	ABR35697
17	65	12.5	757	20	AAV01939
18	65	12.5	757	20	AAV28603
19	65	12.5	757	21	AAV67096
20	65	12.5	911	22	AAV68824
21	64.5	12.4	98	23	ABP07702
22	64.5	12.4	290	23	AAV81832
23	64.5	12.4	445	23	ABP38662
24	64.5	12.4	632	23	ABP28314
25	64	12.3	98	23	ABR10014
26	64	12.3	98	23	ABR64625
27	64	12.3	333	23	ABR93211
28	64	12.3	454	17	AAW04726
29	64	12.3	757	19	AAW71297
30	64	12.3	770	13	AAV26340
31	63.5	12.2	256	13	AAV24796
32	63.5	12.2	378	22	ABG02846
33	63.5	12.2	642	22	ABR61316
34	63.5	12.2	2380	21	AAV18315
35	63.5	12.2	2787	21	AAV81807
36	63	12.1	272	21	AAV14013
37	63	12.1	306	21	AAV14012
38	63	12.1	386	21	AAV14011
39	63	12.1	392	22	AAU37162
40	63	12.1	392	22	AAU37524
41	63	12.1	414	22	AAU34284
42	62.5	12.0	330	22	AAU00443
43	62	11.9	141	21	AAU31554
44	62	11.9	367	22	AAU5083
45	62	11.9	444	23	AAV76669

ALIGNMENTS

RESULT 1	
AAV94526	
AAV94526 standard; protein; 95 AA.	
AC	AAV94526:
XX	
XX	
06-OCT-2000 (first entry)	
XX	
XX	
DE	Human lysine-rich statherin protein.
XX	
XX	
KW	Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
KW	precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
KW	asthma; allergy; diabetes mellitus; fungal; bacterial infection;
KW	cancer; leukemia; adenocarcinoma; melanoma.
XX	
OS	Homo sapiens.
XX	
XX	
Key	Location/Qualifiers
FT	Modified-site 15
FT	/note="Potential casein kinase II phosphorylation site"
FT	Modified-site 24
FT	/note="Potential N-glycosylation site"
FT	Modified-site 29
FT	/note="Potential protein kinase C phosphorylation site"
FT	Modified-site 76
FT	/note="Potential N-glycosylation site"
XX	
XX	
PN	WO200024779-A1.
XX	
XX	
04-MAY-2000.	

Listeria monocytog
Drosophila melanog
Drosophila melanog
Streptococcus poly
Human ORF protein
Fungal ZBC protein
Ecdysone receptor
Ecr polypeptide su
L. cuprina ecdysone
Amino acid sequenc
Human ORF protein
S. epidermidis ope
Staphylococcus epi
Streptococcus poly
Human protein sequ
Human albumin fusl
Herbicideally activ
Aromatic acyl tran
Lucilia cuprina ec
Apt770. Homo sapi
Sequence of fibron
Novel human diagno
Drosophila melanog
Plasmodium falcipa
Human manogany pro
Arabidopsis thalia
Arabidopsis thalia
Staphylococcus aur
Staphylococcus aur
Staphylococcus resl
Maize disease resi
Arabidopsis thalia
Enterococcus faeca
Helicobacter pylori

```

PF 22-OCT-1999; 99WO-US24046.
XX
XX 23-OCT-1998; 98US-0155209.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Corley NC, Guegler KJ, Patterson C;
XX WPI: 2000-350699/30.
XX N-PSDB: AAA48963.
XX
XX Purified polypeptide used for treating or preventing a disorder
XX characterized by expression or activity of lysine-rich statherin
XX proteins -
XX
XX Claim 1; Page 69; 75pp; English.
XX
XX The present sequence is human lysine-rich statherin protein (LRSP). The
XX cDNA sequence encoding this protein was identified through analysis of
XX a cDNA library of breast tumor tissue (BRSTNOT14). The LRSP sequence
XX was found to have homology with human statherin (AA94527) and human
XX basic histidine-rich protein (AA94528). Human statherin is a
XX phosphoprotein that acts as an inhibitor of precipitation of calcium
XX phosphate salts in the oral cavity. The LRSP polypeptide and its
XX antagonists may be useful for treating or preventing disorders
XX associated with the activity of LRSP. Such disorders include
XX autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
XX diabetes mellitus), bacterial and fungal infection and cancers (such
XX as leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be
XX useful for diagnosis of the above disorders.
XX
XX Sequence 95 AA:
SQ
XX
XX Query Match 100.0%; Score 522; DB 21; Length 95;
XX Best Local Similarity 100.0%; Pred. No. 1.8e-57;
XX Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MMFKVGRKQHFVKYFWETDLSNNKTLVSLKKKPFHLYCYIYPLVPKLIILFDIAFI 60
XX 1 MMFKVGRKQHFVKYFWETDLSNNKTLVSLKKKPFHLYCYIYPLVPKLIILFDIAFI 60
XX
XX 61 PKSLISOFONNHYTHNHTNNTNIRFNISNCR 95
XX 61 PKSLISOFONNHYTHNHTNNTNIRFNISNCR 95
XX
XX RESULT 2
XX AAU25643
XX ID AAU25643 standard; Protein; 234 AA.
XX
XX AAU25643;
XX
XX 18-DEC-2001 (first entry)
XX
XX G protein-coupled receptor, ngPCR-2057.
XX
XX Human; mental disorder; thyroid disease; renal failure; anorexia;
XX inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
XX autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
XX depression; Parkinson's disease; Alzheimer's disease; viral infection;
XX Huntington's disease; human immunodeficiency virus; type 2 diabetes;
XX anorexia; hypertension; thrombosis; myocardial infarction;
XX atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
XX ngPCR.
XX
XX Homo sapiens.
XX
XX WO200162924-A2.
XX
XX 30-AUG-2001.
XX
XX 23-FEB-2001; 2001WO-US05989.
XX

```

```

PR 24-FEB-2000; 2000US-0184602.
PR 24-FEB-2000; 2000US-0184604.
PR 24-FEB-2000; 2000US-0184606.
PR 24-FEB-2000; 2000US-0184609.
PR 24-FEB-2000; 2000US-0184650.
PR 24-FEB-2000; 2000US-0184710.
PR 24-FEB-2000; 2000US-0184712.
PR 24-FEB-2000; 2000US-0184715.
PR 24-FEB-2000; 2000US-0184716.
PR 24-FEB-2000; 2000US-0184725.
PR 24-FEB-2000; 2000US-0184822.
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX Vogell G, Wood LS, Parodi LA, Lind P;
XX
XX WPI: 2001-570632/64.
XX N-PSDB: AAS42953.
XX
XX Novel nucleic acid and encoded ngPCR-x, used to screen for compounds
XX for use in the treatment of mental disorders, such as Alzheimer's
XX disease, or Parkinson's disease -
XX
XX Claim 31; Page 69; 263pp; English.
XX
XX The invention relates to novel isolated human G protein-coupled
XX receptors (ngPCR-x). The ngPCR-x can be used for screening compounds
XX which can be used to treat mental disorders, thyroid disease, renal
XX failure, inflammatory conditions such as Crohn's disease, rheumatoid
XX arthritis, autoimmune disorders, schizophrenia, migraine, stroke,
XX dementia, depression, Parkinson's disease, Alzheimer's disease, and
XX Huntington's disease. They may also be used for treating viral infections
XX such as human immunodeficiency virus (HIV), type 2 diabetes, obesity,
XX anorexia, hypotension, hypertension, thrombosis, myocardial infarction,
XX atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726
XX represent the amino acid sequences of novel human G protein-coupled
XX receptors, ngPCR-2031 to ngPCR-2140 respectively, as described in the
XX invention.
XX
XX Sequence 234 AA:
SQ
XX
XX Query Match 14.1%; Score 73.5; DB 22; Length 234;
XX Best Local Similarity 22.1%; Pred. No. 0.61;
XX Matches 21; Conservative 22; Mismatches 37; Indels 15; Gaps 3;
XX
XX 12 EKVTFWETDLSNNKTLVSLKKKPFHL---YCVIYI-PLVPKLIILFDIAFIKSLISQ 67
XX 12 EKVTFWETDLSNNKTLVSLKKKPFHL---YCVIYI-PLVPKLIILFDIAFIKSLISQ 67
XX
XX 133 FSNVLLKPGISNSVSLNAEKKRTIFLPSVCIIFFVLCILPFSVFLFSPHIKNCYVSP 192
XX 133 FSNVLLKPGISNSVSLNAEKKRTIFLPSVCIIFFVLCILPFSVFLFSPHIKNCYVSP 192
XX
XX 68 -----FONNHYTHNHTNNTNIRFNIS 91
XX 68 -----FONNHYTHNHTNNTNIRFNIS 91
XX
XX 193 LLSLNPIMLWFKHRIHAIEAHGEPOVOYCLIS 227
XX 193 LLSLNPIMLWFKHRIHAIEAHGEPOVOYCLIS 227
XX
XX RESULT 3.
XX AAU37060
XX ID AAU37060 standard; Protein; 310 AA.
XX
XX AAU37060;
XX
XX 07-OCT-1999 (first entry)
XX
XX Chlamydia trachomatis lipoprotein sequence.
XX
XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX paratrachoma; inclusion conjunctivitis; genital disease; peritrapatitis;
XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX bartolinitis; pneumonia; venereal lymphogranulomatosis.
XX
XX Chlamydia trachomatis.
XX
XX WO9928475-A2.
XX
XX
XX

```

```

PD 10-JUN-1999.
PF 27-NOV-1998; 98MO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97ER-0015041.
PR 17-DEC-1997; 97ER-0016034.
XX
PA (GEST ) GENSET.
XX
PI Griffiths R;
XX
DR WPI: 1999-371125/31.
XX
PT
XX
Genome sequence of Chlamydia trachomatis
XX
PS Disclosure: Page 877-878; 1755pp; English.
XX
XX
AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nongonemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perihepatitis, bartolinitis; pneumonia in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SQ Sequence 310 AA;
XX
Query Match 14.0%; Score 73; DB 20; Length 310;
Best Local Similarity 26.5%; Pred. No. 0.98;
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps
QY 2 WFH-KVGR-----KQHFKYVFWETDLSNNKTVLSKKKPFHL-----YCVIYIPV 47
Db 118 WEHTKQAGRYSLFEKLSFRAS-----SSSEILREKEPEQQLAIIASPFAYRREN 171
QY 48 PKLILFPIFDIA-FIRKSLIS----QFONNHYYHNHTNHNNTNIRNIIISNCR 95
Db 172 P-----FLSSGFEMKRYVGOTLVLRKNPYYDHAHVELMSIDRFIIPNIT 219
RESULT 4
ID ID ABB94254 standard; Protein; 412 AA.
XX
AC ABB94254;
XX
XX
DT 05-JUN-2002 (first entry)
XX
DE Chlamydia protein sequence SEQ ID NO:362.
XX
XX
KW Chlamydial infection; Chlamydia; vaccine; detection; diagnosis;
KW antigen; antibacterial; immunostimulant; immune response;
KW Chlamydia-specific T-cell response.
XX
XX
OS Chlamydia sp.
XX
XX
MO200208267-A2.
XX
PN
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US23121.
XX
PR 20-JUL-2000; 2000US-0620412.
XX
PR 23-APR-2001; 2001US-0841132.
XX
PA (CORI-) CORIXA CORP.
XX
PI Filing SP, Skelky YAW, Probst P, Bhatia A;

```

	X	XX	WP1; 2002-179901/23.
DR	P	XX	
PT	P	XX	Novel compositions comprising Chlamydia CapI protein and its use in the treatment of Chlamydia infection -
PP	P	XX	
PS	P	XX	Claim 24; Page 322-323; 537pp; English.
CC	C	CC	The present invention describes compositions comprising a Chlamydia CapI protein and methods for the diagnosis and therapy of Chlamydia infection.
CC	C	CC	Chlamydia DNA and protein sequences from the present invention can have antibacterial and immunostimulant activities, and can be used in vaccines.
CC	C	CC	Compounds from the present invention can be used for eliciting an immune response, specifically stimulating a Chlamydia-specific T-cell response or inhibiting the development of a Chlamydia infection in an animal.
CC	C	CC	Methods from the present invention can be used for detecting the presence of Chlamydia in a patient; to stimulate and/or expand T cells specific for a Chlamydia protein; and for treatment of a Chlamydia infection. AB192294 to AB192709 and ABB94096 to ABB94374 represent sequences used in the exemplification of the present invention.
CC	C	CC	
SQ	S	XX	Sequence 412 AA;
			Query Match 14.0%; Score 73; DB 23; Length 412; Best Local Similarity 26.5%; Pred. No. 1.4; Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7.
OY	2	WFH-KVGR-----KQHFKVTFWETDLSNNKTSLVLKKKPFLH-----YCVIIPLV 47 : : : : : : : : : : 95 WHTQAQRYSLLFEKLSEFRAS-----SSSELLELKEPEPOLAILASPFAVRPEN 148	
OY	48	PKLIILFEDIA-FIPKSLS----QFONNHYYTHNTNINTNNIRNPISNCRT 95 : : : : : : : : : : Db 149 P-----FLSSGPFMKTYVGCTLVLRKNPPIYDHAVELHSIDRITLPNIYT 196	
RESULT 5			
ID	ABB94350		ABBS94350 standard; Protein: 426 AA.
AC	ABB94350;		
DT	05-JUN-2002	(first entry)	
DE	Chlamydia trachomatis protein sequence SEQ ID NO:573.		
KW	Chlamydial infection; Chlamydia; vaccine; detection; diagnosis;		
RW	antigen; antibacterial; immunostimulant; immune response;		
XX	Chlamydia-specific T-cell response.		
OS	Chlamydia trachomatis.		
PN	MO200208267-A2.		
PD	31-JAN-2002.		
PR	20-JUL-2001; 2001WO-US23121.		
PT	20-JUL-2000; 2000US-O620412.		
PS	23-APR-2001; 2001US-0841132. (CORI-) CORIXA CORP. Flitng SP, Skeiky YAW, Probat P, Bhatia A; WP1; 2002-179901/23. Novel compositions comprising Chlamydia CapI protein and its use in the treatment of Chlamydia infection - Claim 34; Page 507-508; 537pp; English.		
CS	The present invention describes compositions comprising a Chlamydia CapI		

CC Protein and methods for the diagnosis and therapy of Chlamydia infection.
 CC Chlamydia DNA and protein sequences from the present invention can have
 CC antibacterial and immunostimulant activities, and can be used in
 CC vaccines. Compounds from the present invention can be used for eliciting
 CC an immune response, specifically stimulating a Chlamydia-specific T-cell
 CC response or inhibiting the development of a Chlamydia infection in an
 CC animal. Methods from the present invention can be used for detecting the
 CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
 CC specific for a Chlamydia protein; and for treatment of a Chlamydia
 CC infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent
 CC sequences used in the exemplification of the present invention.

CC Sequence 426 AA;

Query Match 14.0%; Score 73; DB 23; Length 426;
 Best Local Similarity 26.5%; Pred. No. 1.4;
 Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

OY 2 WFR--KVGR-----KQHFVTFWETDLSNNKTVSLKKKPFHL-----YCVIYIPLV 47
 DB 109 WEHTKQAGRYSLFEKLSFRAS-----SSSEILTEKEPEQOLALIASPFVAYRREN 162
 OY 48 PKLILFLDLIA-FIPKSLIS----OFQNNHTYHNTNHTNNTNIRFNISNCR 95
 DB 163 P-----FLSSGPFMPKTYVGQTLVLQKNPYYYDHAHVELHSIDRRIIPNIYT 210

RESULT 6

ABB94255 ID ABB94255 standard; Protein: 433 AA.

XX ABB94255;

DT 05-JUN-2002 (first entry)

DE Chlamydia protein sequence SEQ ID NO:363.

KW Chlamydia infection; Chlamydia; vaccine; detection; diagnosis;
 KW antigen; antibacterial; immunostimulant; immune response;

KW Chlamydia-specific T-cell response.

OS Chlamydia sp.

PN W0200208267-A2.

XX 31-JAN-2002.

PF 20-JUL-2001; 2001WO-US23121.

PR 20-JUL-2000; 2000US-0620412.

PR 23-APR-2001; 2001US-0841132.

PA (CORI-) CORIXA CORP.

PI Filing SP, Skelky YAW, Probst P, Bhatia A;

DR WPI: 2002-179901/23.

PT Novel compositions comprising Chlamydia Cap1 protein and its use in the
 PT treatment of Chlamydia infection -

PS Claim 24; Page 323-324; 537pp; English.

CC The present invention describes compositions comprising a Chlamydia Cap1
 CC protein and methods for the diagnosis and therapy of Chlamydia infection.
 CC Chlamydia DNA and protein sequences from the present invention can have
 CC antibacterial and immunostimulant activities, and can be used in
 CC vaccines. Compounds from the present invention can be used for eliciting
 CC an immune response, specifically stimulating a Chlamydia-specific T-cell
 CC response or inhibiting the development of a Chlamydia infection in an
 CC animal. Methods from the present invention can be used for detecting the
 CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
 CC specific for a Chlamydia protein; and for treatment of a Chlamydia

CC Infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent
 CC sequences used in the exemplification of the present invention.

CC Sequence 433 AA;

Query Match 14.0%; Score 73; DB 23; Length 433;
 Best Local Similarity 26.5%; Pred. No. 1.5;
 Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

OY 2 WFR--KVGR-----KQHFVTFWETDLSNNKTVSLKKKPFHL-----YCVIYIPLV 47
 DB 116 WEHTKQAGRYSLFEKLSFRAS-----SSSEILTEKEPEQOLALIASPFVAYRREN 169

OY 48 PKLILFLDLIA-FIPKSLIS----OFQNNHTYHNTNHTNNTNIRFNISNCR 95
 DB 170 P-----FLSSGPFMPKTYVGQTLVLQKNPYYYDHAHVELHSIDRRIIPNIYT 217

RESULT 7

AAW13825 ID AAW13825 standard; Protein: 1226 AA.

XX AAW13825;

DT 04-JUN-1997 (first entry)

DE Yeast transcription regulatory factor SRB8.

KW Transcription regulatory factor; suppressor of RNA polymerase B;
 KW SRB8; RNA polymerase II; holoenzyme; SWI/SNF.

OS Saccharomyces cerevisiae.

PN W09708301-A1.

XX 06-MAR-1997.

PF 28-AUG-1996; 96WO-US14192.

PR 26-JAN-1996; 96US-0590399.

PR 31-AUG-1995; 95US-0521872.

PR 11-OCT-1995; 95US-0540804.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Chao DM, Koleske AJ, Thompson CM, Young RA;

DR WPI: 1997-179258/16.

DR N-PDB; AAT59908.

PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II
 PT and one or more regulatory proteins, pref. suppressor of RNA
 PT polymerase B proteins or SWI/SNF proteins

PS Claim 11; Fig 10a-b; 154pp; English.

CC Novel yeast SRB (suppressor of RNA polymerase B) proteins SRB4, SRB5,
 CC SRB6, SRB7, SRB8, SRB9, SRB10 and SRB11 (AAW13821-28) are transcription
 CC regulatory factors that act as positive and negative regulators of
 CC RNA polymerase II activity, and are components of the RNA polymerase
 CC II holoenzyme. They were identified using methods designed to
 CC identify transcription factors involved in RNA polymerase II
 CC C-terminal domain (CTD) function. SRB8 and SRB9 appear to repress
 CC CTD activity. Genomic clones (AAT59904-11) for the SRBs have been obtd.
 CC SRBs can be used to treat diseases resulting from alteration or
 CC deletion of the SRB gene, pref. by gene transfer technology. They
 CC can also be used in in vitro transcription of DNA and to identify
 CC cpds. that modify gene transcription.

CC Sequence 1226 AA;

Query Match 13.6%; Score 71; DB 18; Length 1226;
 Best Local Similarity 27.9%; Pred. No. 8.9;

Query Match	13.3%	Score 69.5	DB 23	Length 1381
Best Local Similarity	32.9%	Pred. NO. 16		
Matches 23	Conservative 8	Mismatches 26	Indels 13	Gaps 2
Qy	30 LKKKKPFHLYCYIYIPVPLKLLIFDLIAFIKPSLISOPONNHYTHNHTNTN----- 83			
Db	16 IRRKK-----YIDMLEYLLILLLEFPTLLYSRKNKIKRYDNDLNNINKNNNTNTNTIITYPK 69			
Qy	84 -NIRENIIISN 92			
Db	70 SNISIKIIEEN 79			
RESULT 9				
AAB18244				
ID	AAB18244 standard; Protein; 1245 AA.			
AC	AAB18244;			
XX				
DT	07-NOV-2000 (first entry)			
XX				
DE	Plasmodium falciparum, chromosome 2 related protein SEQ ID NO:101.			
XX				
KM	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;			
KW	antimalarial; malaria; protozoacide; infection; insecticide.			
XX				
OS	Plasmodium falciparum.			
PN	WO200025728-A2.			
XX				
PD	11-MAY-2000.			
XX				
PF	05-NOV-1999; 99WO-US26796.			
XX				
PR	05-NOV-1998; 98US-0107131.			
XX				
PA	(HOEF/) HOFFMAN S.			
PA	(CARU/) CARUCCI D.			
PA	(GARD/) GARDNER M.			
PA	(VENT/) VENTER J C.			
XX				
PI	Hoffman S, Carucci D, Gardner M, Venter JC;			
XX				
DR	WPI: 2000-365347/31.			
XX				
PT	Proteins encoded by chromosome 2 of the human malarial parasite,			
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the			
PI	diagnosis of P.falciparum infection -			
XX				
BS	Disclosure; Page 234-237; 577Pp; English.			
XX				
CC	The present invention describes proteins and their fragments (I) encoded			
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.			
CC	Also described are: (I) nucleotide sequences (II) encoding (I); and (2)			
CC	vaccines against P. falciparum infection comprising (I) or (II).			
CC	(I) and (II) are useful for the development of vaccines against			
CC	P. falciparum infection. (I) and polyclonal antisera or a monoclonal			
CC	antibody raised to immunogens comprising the sequences of (I), are			
CC	useful in the detection of infection with P. falciparum. Furthermore,			
CC	(I) (especially when they are rifins or secreted or membrane proteins)			
CC	can aid the identification of drugs to treat or prevent P. falciparum			
CC	infection, or they can be used to identify drug resistance in			
CC	P. falciparum. Sequencing of the plasmodium chromosome 2 and the			
CC	subsequent identification of proteins encoded by it will help to expand			
CC	our understanding of parasite biology, a process hampered by the			

CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

Sequence 1245 AA;

Query Match

Best Local Similarity 13.28; Score 69; DB 21; Length 1245;

Matches 25; Conservative 12; Mismatches 33; Indels 20; Gaps 3;

QY 22 SNNKTLVSLKKRPFHLYCIVYIPVPLKLIILFDLAFKSLSCFQNNHYTHN--T 78

DB 1069 SNNKFFIIRKKKKLKKLCLCYIMKSEFPHI---LDEFWMLSCONEIKNTYKLNLEHVIS 1124

QY 79 NHTNTNIRFNIIIS-----NCRT 95

DB 1125 LHNSSIIDFKIINHFLNKIFENISINCT 1154

RESULT 10

AAW98706

ID AAW98706 standard; Protein; 346 AA.

AC AAW98706;

DT 31-MAR-1999 (first entry)

DE H. pylori GHP0 727 protein.

XX

KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX peptic ulcer disease.

OS Helicobacter pylori.

XX

PN WO9843478-A1.

PD 08-OCT-1998.

XX

PF 01-APR-1998; 98WO-US06371.

XX

PR 29-JUL-1997; 97US-0902615.

XX

PR 01-APR-1997; 97US-083457.

XX

PR 24-JUN-1997; 97US-0881227.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX

PI A1-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;

DR WPI: 1998-542293/46.

XX

DR N-PSDB; AAX14425.

XX

PT New isolated Helicobacter polynucleotides - used to develop products

PT for the diagnosis, prevention and treatment of Helicobacter

PT infections and gastrointestinal diseases

PS Claim 8; Page 1482-1484; 2054pp; English.

XX

CC This sequence represents a Helicobacter pylori GHP0 protein of the

CC invention. The polypeptides can be used for preventing or treating

CC Helicobacter infections, and gastroduodenal diseases associated with

CC these infections, including acute, chronic, and atrophic gastritis, and

CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be

CC used for the production of antibodies. The products can also be used for

CC detection and diagnosis.

XX

Sequence 346 AA;

Query Match

13.08; Score 68; DB 19; Length 346;

Best Local Similarity 26.38; Pred. No. 4.7;

Matches 26; Conservative 12; Mismatches 45; Indels 16; Gaps 3;

QY 11 HFVY-----TFWETDLSNNKTLVSLKKRPFHLYCIVYIPVPLKLIIL-----FLDIAR 59

DB 146 HFVYKSYSTYVAVSGAGNKGLESLSKNEKLTALCELEKPTIDLNQVLAQAFAPYIAR 205

QY 60 IPRSLISQFQNNHYTHN-----HTNNTNIRFNIIISNC 93

DB 206 NAIAMHIDTFRENGCYTKELKMLHETKIMGVDPIASATC 244

RESULT 11

ABBA9049

ID ABBA9049 standard; Protein; 907 AA.

AC ABBA9049;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #1753.

XX

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

XX vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.

PN WO200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

XX

PR 11-APR-2000; 2000FR-0004629.

XX

PA (INSP) INST PASTEUR.

XX

PI Buchrieser C, Frangeul L, Couve E, Rusnok C, Fsihi H, Delhoux P;

PI Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Kreft J, Kunz M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tlerriz-Martinez A, Amend A;

PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

XX Rose M, Voss H;

XX

DR WPI: 2002-010914/01.

XX

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

PT and prevention of Listeria and related bacterial infections, and

PT related polypeptides

PS Claim 6; SEQ ID No 1754; 192pp; French.

XX

CC The present invention relates to the genome sequence of Listeria

CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of

CC it are useful for selecting probes and primers for detecting genes in L.

CC monocytogenes and related organisms, and for studying genetic

CC polymorphisms and other genomes. The present sequence is a protein

CC encoded by the genome sequence of the present invention. Proteins

CC expressed from the genome sequence are useful for raising specific

CC antibodies, identification of L. monocytogenes and related organisms,

CC and for biosynthesis and biodegradation, especially biosynthesis of vitamin

CC B12. The genome sequence and proteins encoded by it are also useful for

CC selecting compounds that regulate gene expression and cell replication

CC and modulate L. monocytogenes-related diseases. In addition, the genome

CC sequence and proteins encoded by it are useful in pharmaceutical and

CC vaccines compositions for the treatment or prevention of infections by L.

CC monocytogenes and related organisms.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pat_sequences.

XX

20541pa-85

ID	ABB62708 standard; Protein; 1028 AA.
XX	
AC	ABB62708;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 14916.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
PF	23-MAR-2001; 2001WO-US092331.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	MP1; 2001-656860/75.
XX	
N-PSDB	ABU06811.
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX	
PS	Disclosure; SEQ ID NO 14916; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABJ16176-ABL30511), expressed DNA sequences (AAB57737-ABB72072) and the encoded proteins
CC	(AAB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
SO	Sequence 1028 AA;
Query Match	12.8%; Score 67; DB 22; Length 1028;
Best Local Similarity	52.6%; Pred. NO. 23;
Matches 10; Conservative	4; Mismatches 5; Indels 0; Gaps 0;
QY	66 SQFQNNHYTHNHTNTNN 84 : : : 141 SKFHNNHHHQQNNNNNN 159
Db	
RESULT 14	
ABP27420	
ID	ABP27420 standard; Protein; 244 AA.
XX	
AC	ABP27420;
XX	
DT	02-JUL-2002 (first entry)
XX	
DE	Streptococcus polypeptide SEQ ID NO 4016.
XX	
KM	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae; group A streptococcus; Streptococcus pyogenes; antibacterial; antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX	
OS	Streptococcus agalactiae.

XX WO200234771-A2.
 XX 02-MAY-2002.
 PD 29-OCT-2001; 2001WO-GB04789.
 XX 27-OCT-2000; 2000GB-0026333.
 XX 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INSR GENOMIC RES.
 XX Telford J, Maignan V, Margalit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX WPI; 2002-352536/38.
 DR N-PSDB; ABN68051.
 XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX Claim 1; Page 3557; 4525pp; English.
 PS The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GBS (Streptococcus agalactiae) or group A Streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (1), nucleic acids encoding (1), ABN66044-ABN71536 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC Sequence 244 AA;
 SO
 Query Match 12.6%; Score 66; DB 23; Length 244;
 Best Local Similarity 26.0%; Pred. No. 5.5;
 Matches 34; Conservative 14; Mismatches 41; Indels 42; Gaps 8;
 OY 2 WFHVGV-----RKQHF-----KVTFWEDLSNNK-----TLVSEKKR 34
 DB 107 WLTGIGSKRKTHFFELCAHLEVKNSMPPIVILTTDFINNETLKDYIIQSISTDK 166
 OY 35 -PEHLVCVI-----YIPLPKLIIT-----FLDIAFIPKSLISOFONNHTHTNHN 83
 DB 167 IRFHSYLLDNIQISQISTLKPDLITNRKLF---TYTTELATNSLVAAHDVDPQSIS 223
 OY 84 NIRENITSNCR 94
 DB 224 NIO-TTISNIK 233

KW Human: open reading frame; ORFX: gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 OS Homo sapiens.
 XX WO200192523-A2.
 XX 06-DEC-2001.
 PD 29-MAY-2001; 2001WO-US10836.
 PF 30-MAY-2000; 2000US-206132P.
 PR 29-AUG-2000; 2000US-228716P.
 XX (CURA-) CURAGEN CORP.
 PA Shinkets RA, Leach MD;
 PI WPI; 2002-106308/14.
 DR N-PSDB; ABN22752.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX Disclosure; SEQ ID 13982; 1037pp; English.
 PS The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration in various tissues and conditions resulting from
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SO Sequence 61 AA;
 Query Match 12.5%; Score 65.5; DB 23; Length 61;
 Best Local Similarity 34.0%; Pred. No. 1.2;
 Matches 16; Conservative 7; Mismatches 23; Indels 1; Gaps 1;
 OY 37 HLXCVIIPVLPKLIIFLDIAFIPKSLISOFONNHTHTNHN 83
 DB 3 NLXCVTLGIKPSVSYTLVGAQAIIPKSK-DPLQOHQTHNATQSLNS 48

Search completed: April 14, 2003, 22:59:49
 Job time: 111 secs

Tue Apr 15 15:15:18 2003

us-09-830-244b-1.ra1

Page. 1

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 23:01:31 ; Search time 90 Seconds
(without alignments)
31.058 Million cell updates/sec

Title: US-09-830-244B-1
Perfect score: 522
Sequence: 1 MWPHKGRKQHFVPEWED.....NHTNHTNIRNINISNCT 95

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PT05.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	65	12.5	402	2	US-08-290-731C-14
2	65	12.5	911	4	US-09-356-952-6
3	64.5	12.4	445	4	US-09-134-001C-3507
4	63.5	12.2	2787	4	US-09-245-041-15
5	61.5	11.8	863	4	US-09-619-353-14
6	61	11.7	110	2	US-08-569-166-34
7	61	11.7	1226	2	US-08-540-804-12
8	61	11.7	1226	3	US-08-218-265-12
9	61	11.7	1226	3	US-08-521-872-12
10	61	11.7	1226	4	US-08-590-399-12
11	60.5	11.6	166	4	US-09-615-192A-273
12	60.5	11.6	332	2	US-08-671-978A-8
13	60	11.5	429	4	US-08-795-876-44
14	60	11.5	474	1	US-08-453-742-2
15	60	11.5	474	1	US-08-453-742-2
16	60	11.5	474	1	US-08-453-742-2
17	60	11.5	474	1	US-08-453-742-2
18	60	11.5	477	1	US-08-453-742-27
19	60	11.5	477	1	US-08-454-464-27
20	60	11.5	477	1	US-08-453-222-27
21	60	11.5	477	1	US-08-453-222-27
22	59.5	11.4	485	1	US-09-073-297-26
23	59.5	11.4	485	1	US-07-991-867B-42
24	59.5	11.4	485	1	US-08-544-332-42
25	59.5	11.4	485	4	US-09-370-861A-42
26	59.5	11.4	648	4	US-09-370-861A-75
27	59	11.3	99	2	US-08-422-333-3

28	59	11.3	695	1	US-08-339-152A-30	Sequence 30, Appl
29	59	11.3	695	4	US-09-458-481B-4	Sequence 4, Appl
30	59	11.3	695	4	US-09-458-481B-5	Sequence 5, Appl
31	59	11.3	695	4	US-09-458-481B-6	Sequence 6, Appl
32	59	11.3	1311	1	US-08-340-011-5	Sequence 5, Appl
33	59	11.3	2391	3	US-08-901-710-5	Sequence 2, Appl
34	59	11.3	2391	4	US-08-446-855A-2	Sequence 2, Appl
35	59	11.3	2391	4	US-09-150-741-2	Sequence 2, Appl
36	59	11.3	3135	1	US-08-323-170B-2	Sequence 2, Appl
37	59	11.3	3135	4	US-08-954-441-2	Sequence 2, Appl
38	58.5	11.2	281	4	US-09-134-001C-4763	Sequence 4763, Ap
39	58.5	11.2	484	4	US-09-134-001C-5063	Sequence 5063, Ap
40	58.5	11.2	735	2	US-08-765-243-6	Sequence 6, Appl
41	58.5	11.2	735	5	PCT-US95-07285-6	Sequence 6, Appl
42	58.5	11.2	921	4	US-09-206-800-9	Sequence 9, Appl
43	58.5	11.2	1164	2	US-08-589-756-1	Sequence 1, Appl
44	58.5	11.2	1164	4	US-09-206-800-1	Sequence 1, Appl
45	58.5	11.2	1164	4	US-09-206-898-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-290-731C-14
; Sequence 14, Application US/08290731C
; Patent No. 5843646
GENERAL INFORMATION:
APPLICANT: BOWTELL, David Douglas Lawrence
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
TITLE OF INVENTION: SON OF SEVENLESS (MOS) GENE,
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPHEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068
FILING DATE: 17-FEB-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT0921/92
FILING DATE: 17-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: Q-36066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-290-731C-14
Query Match 12.5% Score 65; DB 2; Length 402;
Best Local Similarity 25.2% Pred. No. 3.8;

Matches 28; Conservative 10; Mismatches 23; Indels 50; Gaps 5;

QY 23 NNTLVSL-----KKKPPHLY-----CVIYIPVPLKILFLDLAIFPK-----62
 Db 284 NSKTLASFELNNLTARKNFSNYRDCLENCY--LPCVPLGVYFTDLFLTKGKDNFQ 341

QY 63 -----SLISQFNNHYTHNHTN-----HNTNNI 85
 Db 342 NMINEFKRTVTRILNEIKKFSQVGMFNPINEVOELLNEVISRENTNNI 392

RESULT 2
 US-09-356-952-6
 ; Sequence 6, Application US/09356952
 ; Patent No. 6117663
 ; GENERAL INFORMATION:
 ; APPLICANT: Borjack-Sjodin, Ann
 ; APPLICANT: Margalit, S. M.
 ; APPLICANT: Bor-Sogil, Dafna
 ; APPLICANT: Cole, Philip
 ; APPLICANT: Kurliyan, John
 ; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
 ; FILE REFERENCE: 600-1-228N
 ; CURRENT APPLICATION NUMBER: US/09/356,952
 ; EARLIER FILING DATE: 1999-07-19
 ; EARLIER APPLICATION NUMBER: 60/093,631
 ; EARLIER FILING DATE: 1998-07-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 911
 ; TYPE: PRT
 ; ORGANISM: Schizosaccharomyces pombe
 US-09-356-952-6

Query Match Best Local Similarity 12.5%; Score 65; DB 3; Length 911;
 Matches 28; Conservative 10; Mismatches 23; Indels 50; Gaps 5;

QY 23 NNTLVSL-----KKKPPHLY-----CVIYIPVPLKILFLDLAIFPK-----62
 Db 776 NSKTLASFELNNLTARKNFSNYRDCLENCY--LPCVPLGVYFTDLFLTKGKDNFQ 833

QY 63 -----SLISQFNNHYTHNHTN-----HNTNNI 85
 Db 834 NMINEFKRTVTRILNEIKKFSQVGMFNPINEVOELLNEVISRENTNNI 884

RESULT 3
 US-09-134-001C-3507
 ; Sequence 3507, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3507
 ; LENGTH: 445
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3507

Query Match Best Local Similarity 12.4%; Score 64.5; DB 4; Length 445;
 Best Local Similarity 25.0%; Pred. No. 5;

Matches 21; Conservative 18; Mismatches 26; Indels 19; Gaps 4;

QY 12 FKVTMETDIS-----NNTLVSLKKKKPPHLYCVIYIPVPLKILFLDLAIFPKSLI 65
 Db 168 KISYMLTNTAIVFAGALLGILMGAKRSMLEFIAFYIYIMVFALIV-----WLPKDL- 220

QY 66 SDFNNHYTHNHTNNTNNIRFNI 89
 Db 221 -----NIVTOSHT-HHANEKFSM 238

RESULT 4
 US-09-245-041-15
 ; Sequence 15, Application US/09245041
 ; Patent No. 6274339
 ; GENERAL INFORMATION:
 ; APPLICANT: Moore, K.
 ; APPLICANT: Nagle, D.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
 ; FILE REFERENCE: 7853-136
 ; CURRENT APPLICATION NUMBER: US/09/245,041
 ; EARLIER FILING DATE: 1999-02-05
 ; EARLIER APPLICATION NUMBER: 60/093,630
 ; EARLIER FILING DATE: 1998-07-21
 ; EARLIER APPLICATION NUMBER: 60/104,978
 ; EARLIER FILING DATE: 1998-10-20
 ; NUMBER OF SEQ ID NOS: 131
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 15
 ; LENGTH: 2787
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-245-041-15

Query Match Best Local Similarity 12.2%; Score 63.5; DB 4; Length 2787;
 Matches 29; Conservative 11; Mismatches 25; Indels 31; Gaps 6;

QY 5 KVRKQHPKV-----TFWETDL-----SN--NKTLSLKKKKPPHLYCVIYIPVPLKIL 53
 Db 1611 KLGGKGYKCGGQCYFECKLVVVFYSNLNLVLIHLKLSIFSL-----PVPILHL 1663

QY 54 FLDAIFPKSLISQFNNHYTHNHTNNTNNIRFNI 89
 Db 1664 TLPUSHI-----HTQHT-HNTKCLDFKI 1686

RESULT 5
 US-09-619-353-14
 ; Sequence 14, Application US/09619353
 ; Patent No. 6410249
 ; GENERAL INFORMATION:
 ; APPLICANT: Ngai, John
 ; APPLICANT: Speca, David J.
 ; APPLICANT: Lin, David M.
 ; APPLICANT: Isaacoff, Ehud Y.
 ; APPLICANT: Dittman, Andrew H.
 ; APPLICANT: Fan, Jinhong
 ; TITLE OF INVENTION: Ocular Receptors
 ; FILE REFERENCE: B99-038-2
 ; CURRENT APPLICATION NUMBER: US/09/619,353
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 60/144,766
 ; PRIOR FILING DATE: 1999-07-20
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 14
 ; LENGTH: 863
 ; TYPE: PRT
 ; ORGANISM: Brachydanio rerio (zebrafish)
 US-09-619-353-14

Query Match 11.8%; Score 61.5; DB 4; Length 863;
 Best Local Similarity 28.1%; Pred. No. 27;
 Matches 16; Conservative 9; Mismatches 19; Indels 13; Gaps 2;

OY 43 YIPVPLKLIIT-----FDIAFIPKSLISOFONNHYT-----HNTNNTNNTNR 86
 DB 777 YVPAVENVILISNVGILSCHFLPKRSTYIIIFKREHNTKDAPKKNVETAKSAENIK 833

RESULT 6
 US-08-569-166-34

; Sequence 34, Application US/08569166
 ; Patent No. 5830722

; GENERAL INFORMATION:

; APPLICANT: NICOLAS, LUC

; APPLICANT: CHARLES, JEAN-FRANCOIS

; APPLICANT: DELECLUSE, ARMELE

; APPLICANT: BARLOY, FREDERIQUE

; TITLE OF INVENTION: CLOSTRIDIUM BIFERMENTANS DNA FRAGMENT

; TITLE OF INVENTION: BEARING GENES CODING FOR PROTEINS LINKED TO AN

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/569,166

; FILING DATE: 05-JUL-1996

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PCT/FR94/00768

; FILING DATE: 24-JUN-1994

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: FR 93/07795

; FILING DATE: 25-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 660-106-0 PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 110 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-569-166-34

Query Match 11.7%; Score 61; DB 2; Length 110;
 Best Local Similarity 57.9%; Pred. No. 2,4;
 Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 70 NNHTYNNHTNNTNNTNNTN 88
 DB 6 NNNNTNNTNNTNNTNNTN 24

RESULT 7
 US-08-540-804-12

; Sequence 12, Application US/08540804

; Patent No. 5919666

; GENERAL INFORMATION:

; APPLICANT: Young, Richard A.

; APPLICANT: Koleske, Anthony J.

; APPLICANT: Thompson, Craig M.

; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene

; TITLE OF INVENTION: Transcription and Methods of Use Therefor

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Millitia Drive

; CITY: Lexington

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/540,804

; FILING DATE: 11-OCT-1995

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/521,872

; FILING DATE: 21-AUG-1995

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/218,265

; FILING DATE: 25-MAR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Granahan, Patricia

; REGISTRATION NUMBER: 32,227

; REFERENCE/DOCKET NUMBER: WHI94-03A2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-861-6240

; TELEFAX: 617-861-6240

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1226 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-540-804-12

Query Match 11.7%; Score 61; DB 2; Length 1226;
 Best Local Similarity 26.7%; Pred. No. 49;
 Matches 23; Conservative 13; Mismatches 34; Indels 16; Gaps 4;

OY 11 HFKVTFW-----ETDLSNNKTLVSLKKRPFHLCVYIYIPVPLIIFLDIAFIPKSL 64
 DB 82 HILMFWDICIDITNAPPAATITTS-SQKEPF-----LVTKITDMLIKHYIVSSS 132
 OY 65 ISOFONNHTHTNNTNNTNNTN 90
 DB 133 KSMINDENYIINDIKRN-NKIKLIIIL 157

RESULT 8
 US-08-218-265-12

; Sequence 12, Application US/08218265
 ; Patent No. 5922585

; GENERAL INFORMATION:

; APPLICANT: Young, Richard A.

; APPLICANT: Koleske, Anthony J.

; APPLICANT: Thompson, Craig M.

; TITLE OF INVENTION: No. 5822585el Factors Which Modify Gene

; TITLE OF INVENTION: Transcription and Methods of Use Thereof

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Millitia Drive

CITY: Lexington
STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218,265
FILING DATE: 25-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-218-265-12

Query Match 11.7%; Score 61; DB 2; Length 1226;
Best Local Similarity 26.7%; Pred. No. 49;
Matches 23; Conservative 13; Mismatches 34; Indels 16; Gaps 4;

QY 11 HKVYFW-----ETDLSNNKTVSLKKRPFHLVCYIYIPVPLKLIIFLDIAFIPKSL 64
DB 82 HILMFVNDICQIDTNPVATITS-SQKEPFF-----LVTKITDMLHKYIYSSS 132
QY 65 ISQFONNHYTHNHTNNTNIRFNII 90
DB 133 KSMINDENYIINDIKKN-NKIKINIL 157

RESULT 9
US-08-521-872-12
Sequence 12, Application US/08521872
Patent No. 6015682
GENERAL INFORMATION:
APPLICANT: Young, Richard A.
APPLICANT: Koleske, Anthony J.
APPLICANT: Thompson, Craig M.
TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene
TITLE OF INVENTION: Transcription and Methods of Use Therefor
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/521,872
FILING DATE: 31-AUG-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,265
FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-521-872-12

Query Match 11.7%; Score 61; DB 3; Length 1226;
Best Local Similarity 26.7%; Pred. No. 49;
Matches 23; Conservative 13; Mismatches 34; Indels 16; Gaps 4;

QY 11 HKVYFW-----ETDLSNNKTVSLKKRPFHLVCYIYIPVPLKLIIFLDIAFIPKSL 64
DB 82 HILMFVNDICQIDTNPVATITS-SQKEPFF-----LVTKITDMLHKYIYSSS 132
QY 65 ISQFONNHYTHNHTNNTNIRFNII 90
DB 133 KSMINDENYIINDIKKN-NKIKINIL 157

RESULT 10
US-08-590-399-12
Sequence 12, Application US/08590399
Patent No. 6214588
GENERAL INFORMATION:
APPLICANT: Young, Richard A.
APPLICANT: Koleske, Anthony J.
APPLICANT: Thompson, Craig M.
TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene
TITLE OF INVENTION: Transcription and Methods of Use Therefor
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/590,399
FILING DATE: 26-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/540,804
FILING DATE: 11-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/521,872
FILING DATE: 31-AUG-1995
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/218,265
FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03A3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

LENGTH: 1226 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-590-399-12

Query Match
 Best Local Similarity 11.7%; Score 61; DB 4; Length 1226;
 Matches 23; Conservative 13; Mismatches 34; Indels 16; Gaps 4;

QY 11 HFEVTEW-----ETDLSNNKTVLSKKKPFHLYCVIYIPVPKLIILFLDIAFTPKS 64
 DB 82 HILMIFMNDICQIDTNPAAVATTS-SQKEPF-----LVTKITDMLKRYIVSS 132
 QY 65 ISOFONNHYTHNHTNNTNIRPNT 90
 DB 133 KSMINDENYIINDIKRN-NKIKLNL 157

RESULT 11

US-09-615-192A-273
 Sequence 273, Application US/09615192A
 Patent No. 6410718

GENERAL INFORMATION:
 APPLICANT: Blocksberg, Leonard N.
 TITLE OF INVENTION: Materials and Methods for the
 FILE REFERENCE: 11000.1003C4U
 CURRENT APPLICATION NUMBER: US/09/615,192A
 PRIOR FILING DATE: 2000-07-12
 PRIOR APPLICATION NUMBER: US 08/975,316
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: US 08/713,000
 PRIOR FILING DATE: 1996-09-11
 PRIOR APPLICATION NUMBER: US 09/169,789
 PRIOR FILING DATE: 1998-10-09
 NUMBER OF SEQ ID NOS: 405
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 273
 LENGTH: 166
 TYPE: PRT
 ORGANISM: Eucalyptus grandis
 US-09-615-192A-273

Query Match
 Best Local Similarity 11.6%; Score 60.5; DB 4; Length 166;
 Matches 22; Conservative 9; Mismatches 26; Indels 21; Gaps 4;

QY 13 KVTMETDLSNNKTVLSKKKPFHLYCVIYIPVPKLIILFLDIAFTPKS-LISOFONN 71
 DB 30 KELFRA-----TLVPKIKTYPR-----RVPKIAFMELTGKPLAFLMERFPG 75
 QY 72 H-----YTHNHTNNTN 83
 DB 76 HEGLYSIYHSHPFHAH 93

RESULT 12

US-08-671-978A-8
 Sequence 8, Application US/08671978A
 Patent No. 5959093

GENERAL INFORMATION:
 APPLICANT: Salf, Linda J.
 APPLICANT: Parwani, Anil
 APPLICANT: Kim, Wonyong
 APPLICANT: Chang, Keong-OK
 TITLE OF INVENTION: ROTAVIRUS GENES
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CALFEE, HALTER & GRISWOLD
 STREET: 800 SUPERIOR AVENUE, SUITE 1400
 CITY: CLEVELAND

STATE: OHIO
 COUNTRY: USA
 ZIP: 44114

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/671,978A
 FILING DATE:

CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: GOLRIK, MARY E
 REGISTRATION NUMBER: 34,829
 REFERENCE/DOCKET NUMBER: 22727/00133
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (216) 622-8200
 TELEFAX: (216) 241-0816
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 332 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-08-671-978A-8

Query Match
 Best Local Similarity 11.6%; Score 60.5; DB 2; Length 332;
 Matches 24; Conservative 9; Mismatches 22; Indels 13; Gaps 5;

QY 38 LYCVI--YIPVPKLIILFD---IAFIPKSLI-----SOFONNHYTHNHTN---NTNNI 85
 DB 12 ILCLIMTIIIFRKHIMHFLDLSLIAFYISCTIRLSNAPFANDKLYNGVEGVINTNI 71
 QY 86 RFTISNC 93
 DB 72 -FNVESLC 78

RESULT 13

US-08-795-876-44
 Sequence 44, Application US/08795876
 Patent No. 6403305

GENERAL INFORMATION:
 APPLICANT: Gershengorn, Marvin C.
 APPLICANT: Geras-Raska, Elizabeth
 APPLICANT: Nussenzveig, Daniel R.
 TITLE OF INVENTION: STRATEGY TO CLONE DRUGS FOR G PROTEIN
 NUMBER OF SEQUENCES: 44
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
 STREET: Clinton Square, P.O. Box 1051
 CITY: Rochester
 STATE: New York
 COUNTRY: USA
 ZIP: 14603

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/795,876
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: BRAMAN, SUSAN J
 REGISTRATION NUMBER: 34,103
 REFERENCE/DOCKET NUMBER: 19603/1280

TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1606
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-795-876-44

Query Match 11.5%; Score 60; DB 4; Length 429;
Best Local Similarity 20.8%; Pred. No. 18;
Matches 16; Conservative 15; Mismatches 26; Indels 20; Gaps 2;

OY 4 HVGKROHKEVTFW---ETDLSNNKTIVSLKKRPFHLVCVYIPLVPLIILFDIAFI 60
DB 232 HAITRAVYFNDNCWLSVET-----HLIYIHGPVMAALVNVFFLLNI 274
OY 61 PKSLISOFQNNHYTHNH 77
DB 275 VRVLYTKMRETHEASH 291

RESULT 14

US-08-453-742-2
Sequence 2, Application US/08453742
Patent No. 5622839
GENERAL INFORMATION:
APPLICANT: Moore, Emma E
APPLICANT: Sheppard, Paul O
APPLICANT: Kuestner, Rolf E
TITLE OF INVENTION: Human Calcitonin Receptor
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: One Market Plaza, Stewart St. Tower,
STREET: Twentieth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,742
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/100,887
FILING DATE:
APPLICATION NUMBER: US 07/954,804
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 13952-15-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-742-2

Query Match 11.5%; Score 60; DB 1; Length 474;

Best Local Similarity 20.8%; Pred. No. 20;
Matches 16; Conservative 15; Mismatches 26; Indels 20; Gaps 2;
OY 4 HVGKROHKEVTFW---ETDLSNNKTIVSLKKRPFHLVCVYIPLVPLIILFDIAFI 60
DB 277 HAITRAVYFNDNCWLSVET-----HLIYIHGPVMAALVNVFFLLNI 319
OY 61 PKSLISOFQNNHYTHNH 77
DB 320 VRVLYTKMRETHEASH 336

RESULT 15

US-08-454-464-2
Sequence 2, Application US/08454464
Patent No. 5674689
GENERAL INFORMATION:
APPLICANT: Moore, Emma E
APPLICANT: Sheppard, Paul O
APPLICANT: Kuestner, Rolf E
TITLE OF INVENTION: Human Calcitonin Receptor
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: One Market Plaza, Stewart St. Tower,
STREET: Twentieth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,464
FILING DATE: 30-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/100,887
FILING DATE: 02-AUG-1993
APPLICATION NUMBER: US 07/954,804
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 13952-15-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-454-464-2

Query Match 11.5%; Score 60; DB 1; Length 474;
Best Local Similarity 20.8%; Pred. No. 20;
Matches 16; Conservative 15; Mismatches 26; Indels 20; Gaps 2;

OY 4 HVGKROHKEVTFW---ETDLSNNKTIVSLKKRPFHLVCVYIPLVPLIILFDIAFI 60
DB 277 HAITRAVYFNDNCWLSVET-----HLIYIHGPVMAALVNVFFLLNI 319
OY 61 PKSLISOFQNNHYTHNH 77
DB 320 VRVLYTKMRETHEASH 336

Search completed: April 14, 2003, 23:06:48

Tue Apr 15 15:15:18 2003

us-09-830-244b-1.rai

Page 7

Job time : 91 secs

Run on: April 14, 2003, 23:03:36 ; Search time 163 Seconds

(Without alignments)
35.631 Million cell updates/sec

```
Title: US-09-830-24A-B-1
Perfect score: 522
Sequence: 1 MMTFKVGRQHFKYTFWETD.....NNTNNTNNTRENTISNCRT 95
```

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

number of hits satisfying chosen parameters: 248812

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  publishedApplications_AA: *
2:  /cgn2_6/p/odataa/1/pubppa/US08_NEW_PUB_dep: *
3:  /cgn2_6/p/odataa/1/pubppa/PCCT_NEW_PUB_dep: *
4:  /cgn2_6/p/odataa/1/pubppa/US06_NEW_PUB_dep: *
5:  /cgn2_6/p/odataa/1/pubppa/PUBCOMB_dep: *
6:  /cgn2_6/p/odataa/1/pubppa/US07_PUBCOMB_dep: *
7:  /cgn2_6/p/odataa/1/pubppa/PCNUS_PUBCOMB_dep: *
8:  /cgn2_6/p/odataa/1/pubppa/PCNUS_PUBCOMB_dep: *
9:  /cgn2_6/p/odataa/1/pubppa/US08_PUBCOMB_dep: *
10: /cgn2_6/p/odataa/1/pubppa/US09_PUBCOMB_dep: *
11: /cgn2_6/p/odataa/1/pubppa/US10_NEW_PUB_dep: *
12: /cgn2_6/p/odataa/1/pubppa/US10_PUBCOMB_dep: *
13: /cgn2_6/p/odataa/1/pubppa/US60_NEW_PUB_dep: *
14: /cgn2_6/p/odataa/1/pubppa/US60_PUBCOMB_dep: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Query					Description	
No.	Score	Match	Length	ID		
1	73.5	14.1	234	9	US-09-791-379-137	Sequence 137, App
2	73	14.0	412	10	US-09-841-133-362	Sequence 362, App
3	73	14.0	436	10	US-09-841-133-573	Sequence 573, App
4	73	14.0	433	10	US-09-841-133-573	Sequence 573, App
5	66	12.6	2150	9	US-10-135-922-117	Sequence 363, App
6	63.5	12.2	2787	10	US-09-893-238-15	Sequence 17, App
7	63	12.1	332	10	US-09-813-243-12755	Sequence 15, App
8	63	12.1	332	10	US-09-813-243-12755	Sequence 12755, App
9	63	12.1	434	10	US-09-815-242-11117	Sequence 11117, App
10	62	11.9	121	12	US-09-815-242-5780	Sequence 5780, App
11	62	11.9	367	10	US-10-078-929-118	Sequence 118, App
12	62	11.9	1092	9	US-09-815-242-10676	Sequence 10676, App
13	61	11.7	961	10	US-09-823-126-5	Sequence 5, App
14	61	11.7	1427	10	US-09-801-368-132	Sequence 132, App
15	60.5	11.6	448	10	US-09-801-368-354	Sequence 354, App
16	60.5	11.6	467	10	US-09-361-443-4	Sequence 4, App
17	60	11.5	157	10	US-09-361-443-2	Sequence 2, App
18	60	11.5	490	9	US-09-881-7524-200	Sequence 200, App
19	60	11.5	666	10	US-10-051-186-8	Sequence 8, App
					US-09-801-368-36	Sequence 36, App

20	59.5	11.4	117	10	US-09-829-482-26	Sequence 26, Appl
21	58.5	11.2	255	9	US-09-866-050A-725	Sequence 725, Appl
22	58.5	11.2	1164	10	US-09-870-132-1	Sequence 1, Appl
23	58	11.1	99	9	US-10-183-119-2	Sequence 2, Appl
24	58	11.1	100	10	US-09-794-975-4	Sequence 4, Appl
25	58	11.1	103	10	US-09-972-475-2	Sequence 2, Appl
26	58	11.1	103	10	US-09-895-443-2	Sequence 2, Appl
27	58	11.1	117	9	US-09-422-569-10	Sequence 2, Appl
28	58	11.1	117	9	US-09-794-975-6	Sequence 10, Appl
29	58	11.1	117	10	US-09-823-153-2	Sequence 6, Appl
30	58	11.1	355	10	US-09-794-975-13	Sequence 2, Appl
31	58	11.1	695	10	US-09-794-927-10	Sequence 13, Appl
32	58	11.1	695	10	US-09-794-927-12	Sequence 10, Appl
33	58	11.1	695	10	US-09-794-927-11	Sequence 12, Appl
34	58	11.1	695	10	US-09-795-447-10	Sequence 14, Appl
35	58	11.1	695	10	US-09-795-447-11	Sequence 10, Appl
36	58	11.1	695	10	US-09-795-447-12	Sequence 12, Appl
37	58	11.1	695	10	US-09-794-743-10	Sequence 14, Appl
38	58	11.1	695	10	US-09-794-743-11	Sequence 12, Appl
39	58	11.1	695	10	US-09-794-743-12	Sequence 14, Appl
40	58	11.1	695	10	US-09-794-743-14	Sequence 10, Appl
41	58	11.1	695	10	US-09-794-748-10	Sequence 12, Appl
42	58	11.1	695	10	US-09-794-748-12	Sequence 14, Appl
43	58	11.1	695	10	US-09-794-748-14	Sequence 10, Appl
44	58	11.1	695	10	US-09-794-925-10	Sequence 14, Appl
45	58	11.1	695	10	US-09-794-925-12	Sequence 12, Appl
	58	11.1	695	10	US-09-794-925-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-791-279-137

Sequence 137, Application US/09791279
Publication No. US20030050456A1
GENERAL INFORMATION

; GENERAL INFORMATION:

APPLICANT: Vogel, Gabriele
; APPLICANT: Wood, Linda S.
APPLICANT: Wood, Linda S.

```

APPLICANT: Parodi, Luis
; APPLICANT: Lind, Peter

```

FILE REFERENCE: 00048.051
TITLE OF INVENTION: No. US20030050456a1 G Protein-Coupled Receptors

CURRENT APPLICATION NUMBER: US/09/791,279
CURRENT FILING DATE: 2001-03-22

CONTRACT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: 60/184,715
PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/184725

;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: 60/184,712

PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/194 606

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,604
PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,822
PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/184,710
PRIOR FILING DATE: 2000-03-24

PRIOR APPLICATION NUMBER: 60/184,689

PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,690

PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/184,716

PRIOR FILING DATE: 2000-02-24
NUMBER OF SEQ ID NOS. 230

SOFTWARE: PatentIn version 3.0
 SEQ ID NO 137

LENGTH: 234

ORGANISM: Homo sapiens

US-09-791-279-137

Query Match 14.1%; Score 73.5; DB 9; Length 234;
Best Local Similarity 22.1%; Pred. No. 0.45;
Matches 21; Conservative 22; Mismatches 37; Indels 15; Gaps 3;

QY 12 EVYVETDLSNNKTLVSLKRRKPHL-----YCVIYI-PLVPLILFLFDIAFIPKSLISQ 67
DB 133 FEMVLLKPLGSLNSVSYLNKRRKRTITLIPSVCIIFVLGLIRSVYFLSPFHKNQVSP 192
QY 68 -----FQNNHYTHNHTNNTNIRNIIIS 91
DB 193 LLSLNPWLWPKHRRHRIHAIEHGEPOVYCLIS 227

RESULT 2
US-09-841-132-362
Sequence 362, Application US/09841132
Patent No. US20020061848A1
GENERAL INFORMATION:
APPLICANT: Bhalla, Ajay
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
FILE REFERENCE: 210121.469C8
CURRENT APPLICATION NUMBER: US/09/841.132
CURRENT FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 599
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 362
LENGTH: 412
TYPE: PRT
ORGANISM: Chlamydia
US-09-841-132-362

Query Match 14.0%; Score 73; DB 10; Length 412;
Best Local Similarity 26.5%; Pred. No. 1;
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQFKVETWETDLSNNKTLVSLKRRKPHL-----YCVIYIPLV 47
DB 95 WEHTKQAGRYSLFEKLSFRAS-----SSSEILIEKPEPQLAILASPFVAVYRPN 148
QY 48 PKLILFLDIA-FIPKSLIS-----QFQNNHYTHNHTNNTNIRNIIISNCR 95
DB 149 P-----FLSSGPFMPKTYVQGTLLVQKNPYYDHAHVELHSIDFRILPNIYT 196

RESULT 3
US-09-841-132-573
Sequence 573, Application US/09841132
Patent No. US20020061848A1
GENERAL INFORMATION:
APPLICANT: Bhalla, Ajay
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
FILE REFERENCE: 210121.469C8
CURRENT APPLICATION NUMBER: US/09/841.132
CURRENT FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 599
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 573
LENGTH: 426
TYPE: PRT
ORGANISM: C. Trachomatis D serovar
US-09-841-132-573

Query Match 14.0%; Score 73; DB 10; Length 426;
Best Local Similarity 26.5%; Pred. No. 1;
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQFKVETWETDLSNNKTLVSLKRRKPHL-----YCVIYIPLV 47
DB 109 WEHTKQAGRYSLFEKLSFRAS-----SSSEILIEKPEPQLAILASPFVAVYRPN 162
QY 48 PKLILFLDIA-FIPKSLIS-----QFQNNHYTHNHTNNTNIRNIIISNCR 95
DB 163 P-----FLSSGPFMPKTYVQGTLLVQKNPYYDHAHVELHSIDFRILPNIYT 210

RESULT 4
US-09-841-132-363
Sequence 363, Application US/09841132
Patent No. US20020061848A1
GENERAL INFORMATION:
APPLICANT: Bhalla, Ajay
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
FILE REFERENCE: 210121.469C8
CURRENT APPLICATION NUMBER: US/09/841.132
CURRENT FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 599
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 363
LENGTH: 433
TYPE: PRT
ORGANISM: Chlamydia
US-09-841-132-363

Query Match 14.0%; Score 73; DB 10; Length 433;
Best Local Similarity 26.5%; Pred. No. 1.1;
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQFKVETWETDLSNNKTLVSLKRRKPHL-----YCVIYIPLV 47
DB 116 WEHTKQAGRYSLFEKLSFRAS-----SSSEILIEKPEPQLAILASPFVAVYRPN 169
QY 48 PKLILFLDIA-FIPKSLIS-----QFQNNHYTHNHTNNTNIRNIIISNCR 95
DB 170 P-----FLSSGPFMPKTYVQGTLLVQKNPYYDHAHVELHSIDFRILPNIYT 217

RESULT 5
US-10-135-322-17
Sequence 17, Application US/10135322
Patent No. US20020173017A1
GENERAL INFORMATION:
APPLICANT: BENFEY, PN
APPLICANT: HELARIJUTTA, Y
APPLICANT: MAHONEN, AP
APPLICANT: BONKE, AMM
APPLICANT: KAUPPINEN, L
APPLICANT: RIIKONEN, M
TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF
FILE REFERENCE: 5914-0086-999
CURRENT APPLICATION NUMBER: US/10/135.322
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: 60/253,739
NUMBER OF SEQ ID NOS: 43
SOFTWARE: PatentIn version 3.0
SEQ ID NO 17
LENGTH: 2150
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-135-322-17

Query Match 12.6%; Score 66; DB 9; Length 2150;
Best Local Similarity 44.8%; Pred. No. 46;
Matches 13; Conservative 5; Mismatches 11; Indels 0; Gaps 0;
QY 66 SQFQNNHYTHNHTNNTNIRNIIISNCR 94


```
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5780
LENGTH: 414
TYPE: PRT
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(414)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5780

Query Match      12.1% Score 63; DB 10; Length 414;
Best Local Similarity 31.2% Pred. No. 15;
Matches 15; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

QY 39 KCVIYIPVPLKILIFLDIAFIKPSL-ISOFONNHYTHNHTNNTNI 85
DB 192 YTFLEL-ITVSIIVGFLMAYLPKQDEIQGVNHEHPSEHNHTSKI 238

RESULT 10
US-10-078-929-118
Sequence 118, Application US/10078929
Patent No. US20020152497A1
GENERAL INFORMATION:
APPLICANT: Rafelski, Antoni
APPLICANT: Miao, Guo-Hua
APPLICANT: Sakai, Saverio Carl
APPLICANT: Sakai, Hajime
APPLICANT: Famodu, Omolayo O.
APPLICANT: Odell, Joan T.
APPLICANT: Meyers, Blake
APPLICANT: Thorpe, Catherine
APPLICANT: Meng, Zude
TITLE OF INVENTION: Nucleic Acid Fragments Encoding Proteins Involved in
FILE REFERENCE: BBI357 US NA
CURRENT APPLICATION NUMBER: US/10/078,929
PRIOR FILING DATE: 2002-02-19
PRIOR APPLICATION NUMBER: 09/566,394
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: 60/133038
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: 60/133042
PRIOR FILING DATE: 1999-05-07
```

```
PRIOR APPLICATION NUMBER: 60/133427
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/133437
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/133428
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/133438
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/133436
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/137667
PRIOR FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 208
SOFTWARE: Microsoft Office 97
SEQ ID NO 118
LENGTH: 121
TYPE: PRT
ORGANISM: Oryza sativa
US-10-078-929-118

Query Match      11.9% Score 62; DB 12; Length 121;
Best Local Similarity 23.9% Pred. No. 4;7;
Matches 21; Conservative 14; Mismatches 29; Indels 24; Gaps 4;

QY 17 WETDLSNNKTVLSKKR--KPEHL-YCVIYIPVPLP-----KLIIFLDIAFIKPSL 64
DB 37 WELQTDNDNDIMPALKSLDYDLFPFHLQOCFIYCALPEPDYKFPDSELIHMGIGDLI----- 92
QY 65 ISOFONNHYTHNHTNNTNIRFNIIISN 92
DB 93 -----QSHODONKRTEDIALSCINH 112

RESULT 11
US-09-815-242-10676
Sequence 10676, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10676
LENGTH: 367
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10676
```

```

Query Match      11.98; Score 62; DB 10; Length 367;
Best Local Similarity 28.38; Fred No. 17;
Matches 15; Conservative 12; Mismatches 20; Indels 6; Gaps 1;

QY    21  LNNKTTLSAKKKRPHLCVYIIPLVKRLITFDIAFIRKSLSISOFONNHY 73
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     33  IKORWTELELORK-----VFELPILVAIVHVISLTLLTCLLFSLSVQGQGY 79

```

RESULT 12
US-09-423-126-5

```

Sequence 5, Application US/09423126
Patent No. US20020170083A1
GENERAL INFORMATION:
APPLICANT: Buchter-Larsen, et al.
TITLE OF INVENTION: A PROCESS OF PREPARING AN ANTI-OXIDANT
FILE REFERENCE: 674509-2020
CURRENT APPLICATION NUMBER: US/09/423.126
CURRENT FILING DATE: 1999-11-05
PRIOR APPLICATION NUMBER: PCT/IB98/00708
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: GB 9709161.5
PRIOR FILING DATE: 1997-05-06
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 1092
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: fungus sp. or fungus infected graciariopsis sp
US-09-423-126-5

```

Query Match	11.9%	Score 62;	DB 9;	Length 1092;
Best Local Similarity	23.1%	Pred. No. 61;		
Matches 27; Conservative	15;	Mismatches 29;	Totals 46;	Coverage 25%

6 VGRQHFKVTEWETDLSNNKTLVLSKKRP--FLUCVIVPLVPRKILLFLDAIPIKS 6
40 VRKSHYVSM--TALSDDKQTAISIGPDNPGLINQVNDYIPIVA-----GTPPLS 8

```

04 -----LISQFQNNHYTHNH-----TNNHTNN--IREN 88
      : | : : | : : | : : | : : |
88 NTNWYAAGSSTPGITDWTATMNVKEDRIDNPSSYNNHPVQIQVTSYNNNSFRIRN 144

```

RESULT 13
US-09-801-368-132

Sequence 132, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 132
; LENGTH: 961
; TYPE: PR1
; ORGANISM: Saccharomyces cerevisiae
us-09-801-368-132
```

Query Match	11.7%;	Score 61;	DB 10;	Length 961;
Best Local Similarity	43.5%;	Pred. No. 69;		
Matches 10;	Conservative	8;	Mismatches	5; Indels 0; Gaps 0;

```
QY 70 NNHYTHNHTNHTNNIRENIISN 92
    ||: :|: |:|:|: |: :|
Db 533 NNNNNNNNNNNNNNNIKKNVDNN 555
```

RESULT 14
US-09-801

```

Sequence 354 Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fur
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
PRIOR FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 354
LENGTH: 1427
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
ID-09-801-368-354

```

Query Match	11.7%	Score 61;	DB 10;	Length 1427;
Best Local Similarity	26.7%;	Pred. No. 1.1e+02;		
Matches	23;	Conservative	13;	Mismatches 34;
			Indels	16;
			Gaps	4;

```

11 HEKVTEN-----ETDLSNNKLVKLVKKKPEHLYCVIYIPVLPKLIILFLDAIFPKSL 64
   | : || :| : : || || : : | : | : | : | : | : | : | : | : | : | : |
283 HILMIMWMDICQIDNAPVAATITS-SQKEPF-----LVTKIDMLHMKYIYSSS 333

```

334 KSMINDENYIINDIKKN-NKIKLNTL 356

RESULT 15
S-09-361

Sequence 4, Application US/09361443
Patent No. US20020150591A1
GENERAL INFORMATION:
APPLICANT: Muidin, Andrew D
APPLICANT: Oomen, Raymond P
TITLE OF INVENTION: CHLAMYDIA ANTIGENS AND CORRESPONDING DNA FRAGMENTS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 19721-005 (AV-5)

Tue Apr 15 15:15:19 2003

us-09-830-244b-1.rapb

Page 6

```

: CURRENT APPLICATION NUMBER: US/09/361,443
: CURRENT FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: 60/094,195
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 448
: TYPE: PRT
: ORGANISM: Chlamydia pneumoniae
US-09-361-443-4

```

Query Match	11.6%	Score 60.5	DB 10,	Length 448,
Best Local Similarity	29.3%	Pred. No. 32,		
Matches 24; Conservative	14;	Mismatches 23;	Indels 21;	Gaps 5,

Search completed: April 14, 2003, 23:09:45
Job time : 164 secs

•

AL512306/c
 LOCUS AL512306 133984 bp DNA linear PRI 01-FEB-2002
 DEFINITION Human DNA sequence from clone RP11-430C7 on chromosome 1, complete
 sequence.
 ACCESSION AL512306
 VERSION AL512306.16 GI:18491332
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 133984)
 Bray-Allen, S.
 DIRECT SUBMISSION
 Submitted (01-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquer@sanger.ac.uk
 On Feb 4, 2002 this sequence version replaced gi:18477311.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Ch1
 RP11-430C7 is from the library RP11-11.2 constructed by the group
 of Pletier de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-430C7. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true right end of clone RP11-430C7 is at 133984 in this
 sequence. The true left end of clone RP11-739N20 is at 97715 in
 this sequence. The true right end of clone RP11-2317 is at 2000 in
 this sequence.

FEATURES
 source
 1. .133984
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-430C7"
 /clone_lib="RP11-11.2"
 123904..123933
 /note="Single clone region. Reads generated from a
 transposon library derived from a single pUC clone.
 Restriction digest data confirm the assembly."

BASE COUNT 37368 a 32097 c 30202 g 34317 t

ORIGIN
 Alignment Scores:
 Fried, No.: 8,83e-28 Length: 133984
 Score: 349.00 Matches: 65
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 98.48% Mismatches: 0
 Query Match: 66.86% Indels: 0
 DB: 9 Gaps: 0

US-09-830-244b-1 (1-95) x AL512306 (1-133984)
 QY 1 MettPphenHstysValGtlyArgLysGlnHnsPheLysValThrPheTrpGluTrpAsp 20
 |||||||
 Db 91100 ATGTGGTTTCATTAAGGGGAGAGAAAAACAGCATTTTAAAGTAACCTTTTGGGAGACTGAT 91041
 |||||||
 QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysProPheHstLeuTrpCys 40
 |||||||
 Db 91040 TTGAGTAATAATAAAGCTCGCTCCCTTAAGAAAAAACCCTTCACCTTACTGT 90981
 |||||||
 QY 41 ValIleTyrIleProLeuValProLysLeuIleIleLeuPheLysAspIleAlaPheIle 60
 |||||||
 Db 90980 GTCATTATATATCCCTTACTGTCACAAAGTAATATCTTATTTCTGATTTCTTTATA 90921
 |||||||
 QY 61 PolySerLeuIleSer 66
 |||||||
 Db 90920 CCNAACACCTTATTCAGC 90903
 |||||||

RESULT 2
 AC021462 153023 bp DNA linear HTG 03-APR-2000
 LOCUS AC021462
 DEFINITION Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
 pieces.
 AC021462
 AC021462.3 GI:7387343
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 153023)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens, clone RP11-276C1
 Unpublished
 2 (bases 1 to 153023)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barre, N., Beckert, R., Bede, F.,
 Boguski, L., Boukhalil, B., Brown, A., Burkett, G., Castle, A.,
 Choe, X., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 DeArnell, K., Dewar, R., Dominko, M., Doyle, M., Fenebor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lander, T., Lechick, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 MacDonald, P., Marquis, N., McEwen, P., McGurk, A., McKernan, K.,
 McNeeters, R., Melidim, J., Meneau, L., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,
 Pletier, N., Pisan, C., Pollard, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 3, 2000 this sequence version replaced gi:7230200.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5201
 Center clone name: 276.C.1
 ----- Summary Statistics
 Sequencing vector: M13; M77815, 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 141815 bases at least Q40
 Consensus quality: 147368 bases at least Q30

Consensus quality: 149481 bases at least Q20
 Insert size: 15700; agarose-fp
 Quality coverage: 4.3 in Q20 bases; agarose-fp
 Quality coverage: 4.5 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1015: contig of 1015 bp in length
1016 1115: gap of 100 bp
1116 2276: contig of 1161 bp in length
2277 2376: gap of 100 bp
2377 4009: contig of 1633 bp in length
4010 4109: gap of 100 bp
4110 5737: contig of 1628 bp in length
5738 5837: gap of 100 bp
5838 6939: contig of 1102 bp in length
6940 7039: gap of 100 bp
7040 9504: contig of 2465 bp in length
9505 9604: gap of 100 bp
11735 11734: contig of 2130 bp in length
11735 11834: gap of 100 bp
11835 14111: contig of 2277 bp in length
14112 14211: gap of 100 bp
14212 16973: contig of 2762 bp in length
16974 17073: gap of 100 bp
17074 19588: contig of 2515 bp in length
19589 19688: gap of 100 bp
19689 22275: contig of 2587 bp in length
22276 22375: gap of 100 bp
22376 25823: contig of 3448 bp in length
25824 25923: gap of 100 bp
31308 31307: contig of 5384 bp in length
31308 31407: gap of 100 bp
31408 34863: contig of 3456 bp in length
34864 34963: gap of 100 bp
34964 41362: contig of 6419 bp in length
41363 41482: gap of 100 bp
41483 49024: contig of 7542 bp in length
49025 49124: gap of 100 bp
49125 58672: contig of 9548 bp in length
58673 58772: gap of 100 bp
58773 69622: contig of 10850 bp in length
69623 69722: gap of 100 bp
69723 88191: contig of 18469 bp in length
88192 88291: gap of 100 bp
88292 107084: contig of 18793 bp in length
107085 107184: gap of 100 bp
107185 153023: contig of 45839 bp in length.
Location/Qualifiers
1. .153023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCT-11 Human Male BAC"
1. .1015
/misc_feature
/note="assembly-fragment"
1116. .2276
/note="assembly-fragment"
2377. .4009
/note="assembly-fragment"
4110. .5737
/note="assembly-fragment"
5838. .6939
/note="assembly-fragment"
7040. .9504
/note="assembly-fragment"

```

```

misc_feature 9605. .11734
/note="assembly-fragment"
misc_feature 11835. .14111
/note="assembly-fragment"
misc_feature 14212. .16973
/note="assembly-fragment"
misc_feature 17074. .19588
/note="assembly-fragment"
misc_feature 19689. .22275
/note="assembly-fragment"
misc_feature 22376. .25823
/note="assembly-fragment"
misc_feature 25924. .31307
/note="assembly-fragment"
misc_feature 31408. .34863
/note="assembly-fragment"
misc_feature 34964. .41382
/note="assembly-fragment"
misc_feature 41483. .49024
/note="assembly-fragment"
misc_feature 49125. .58672
/note="assembly-fragment"
misc_feature 58773. .69622
/note="assembly-fragment"
misc_feature 69723. .88191
/note="assembly-fragment"
misc_feature 88292. .107084
/note="assembly-fragment"
misc_feature 107185. .153023
/note="assembly-fragment"
misc_feature 107185. .153023
/note="assembly-fragment"

```

BASE COUNT 41494 a 35192 c 34573 g 39756 t 2008 others
 ORIGIN
 vector:side:left"

Alignment Scores:
 Pred. No.: 1.01e-27 Length: 153023
 Score: 349.00 Matches: 65
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 98.48% Mismatches: 0
 Query Match: 66.86% Indels: 0
 DB: 2 Gaps: 0

US-09-830-244b-1 (1-95) x AC021462 (1-153023)

```

QY 1 MetTrpPheHisLysValGlyArgLysGlnHisPheLysValAlaThrPheTrpGluThrAsp 20
|||||
DB 442 ATGTGGTTTCATTAAGTGGAGAGAAAACAGCATTTAAAGTAAGTCTTTGGAGACTGAT 501
|||||
QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysProPheHisLeuTyrcys 40
|||||
DB 502 TTGAGTAATAAATTAACCTCGTCCCTTAAGCAAAAAACCTTCCACCTTTACTGT 561
|||||
QY 41 ValIleTyrIleProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIle 60
|||||
DB 562 GTCATTATATCCCTTCACTTCCCAAGTAATATCTATTCTCGATATGCTTTTATA 621
|||||
QY 61 ProLysSerLeuIleSer 66
|||||
DB 622 CCAAGACCCCTTATCAGC 639
|||||

```

RESULT 3
 AL450424 146805 bp DNA linear HTG 12-SEP-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-561I16.*** SEQUENCING IN
 DEFINITION PROGRAMS *** 25 unordered pieces.
 ACCESSION AL450424 GI:12331147
 VERSION HTG: HTGS-PHASE1; HTGS-CANCELLED.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

1 (bases 1 to 146805)
McLay, K.
Direct Submission
Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced g1:11493361.

Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

Project Information
Center project name: BA563116

Summary Statistics

Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 135323 bases at least Q40
Consensus quality: 139636 bases at least Q30
Insert size: 14405; sum-of-contigs
Insert size: 171032; 2.28 error; agarose-fp
Quality coverage: 3.60x in Q20 bases; sum-of-contigs Quality
Coverage: 3.25x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5105: contig of 5105 bp in length
* 5106 5205: gap of 100 bp
* 5206 10570: contig of 5365 bp in length
* 10571 10670: gap of 100 bp
* 10671 17626: contig of 6956 bp in length
* 17627 17726: gap of 100 bp
* 17727 20478: contig of 2752 bp in length
* 20479 20578: gap of 100 bp
* 20579 23035: contig of 2457 bp in length
* 23036 23135: gap of 100 bp
* 23136 25308: contig of 2173 bp in length
* 25309 25408: gap of 100 bp
* 25409 30469: contig of 5061 bp in length
* 30470 30569: gap of 100 bp
* 30570 32688: contig of 2119 bp in length
* 32689 32788: gap of 100 bp
* 32789 36187: contig of 3399 bp in length
* 36188 36287: gap of 100 bp
* 36288 48988: contig of 12701 bp in length
* 48989 49088: gap of 100 bp
* 49089 67333: contig of 18245 bp in length
* 67334 67433: gap of 100 bp
* 67434 82317: contig of 14884 bp in length
* 82318 82417: gap of 100 bp
* 82418 84614: contig of 2197 bp in length
* 84615 84714: gap of 100 bp
* 84715 90326: contig of 5612 bp in length
* 90327 90426: gap of 100 bp
* 90427 92647: contig of 2221 bp in length
* 92648 92747: gap of 100 bp
* 92748 97690: contig of 4943 bp in length
* 97691 97790: gap of 100 bp
* 97791 100799: contig of 3009 bp in length
* 100800 100899: gap of 100 bp
* 100900 104202: contig of 3303 bp in length
* 104203 104302: gap of 100 bp
* 104303 107157: contig of 2855 bp in length
* 107158 107257: gap of 100 bp
* 107258 110300: contig of 3043 bp in length

FEATURES
source

misc_feature	110301..110400: gap of 100 bp
misc_feature	110401..113955: contig of 3555 bp in length
misc_feature	113956..114055: gap of 100 bp
misc_feature	114056..118801: contig of 4746 bp in length
misc_feature	118802..118901: gap of 100 bp
misc_feature	118902..129831: contig of 10930 bp in length
misc_feature	129832..129931: gap of 100 bp
misc_feature	129932..137545: contig of 7614 bp in length
misc_feature	137546..137645: gap of 100 bp
misc_feature	137646..146805: contig of 9160 bp in length.
misc_feature	Location/Qualifiers
misc_feature	1..146805
misc_feature	/organism="Homo sapiens"
misc_feature	/db_xref="taxon:9606"
misc_feature	/chromosome="1"
misc_feature	/clone_id="RPC1-11.2"
misc_feature	1..5105
misc_feature	/note="assembly-fragment:00368"
misc_feature	clone_end:T7
misc_feature	vector_side:left"
misc_feature	5206..10570
misc_feature	/note="assembly-fragment:00594"
misc_feature	fragment_chain:1"
misc_feature	10671..17626
misc_feature	/note="assembly-fragment:00786"
misc_feature	fragment_chain:1"
misc_feature	17727..20478
misc_feature	/note="assembly-fragment:00844"
misc_feature	fragment_chain:1"
misc_feature	20579..23035
misc_feature	/note="assembly-fragment:00011"
misc_feature	fragment_chain:1"
misc_feature	23136..25308
misc_feature	/note="assembly-fragment:01029"
misc_feature	fragment_chain:1"
misc_feature	25409..30469
misc_feature	/note="assembly-fragment:00148"
misc_feature	fragment_chain:2"
misc_feature	30570..32688
misc_feature	/note="assembly-fragment:00961"
misc_feature	fragment_chain:2"
misc_feature	32789..36187
misc_feature	/note="assembly-fragment:00310"
misc_feature	fragment_chain:3"
misc_feature	36288..48988
misc_feature	/note="assembly-fragment:01150"
misc_feature	fragment_chain:3"
misc_feature	49089..67333
misc_feature	/note="assembly-fragment:00444"
misc_feature	fragment_chain:4"
misc_feature	67434..82317
misc_feature	/note="assembly-fragment:00140"
misc_feature	fragment_chain:4"
misc_feature	82418..84614
misc_feature	/note="assembly-fragment:00657"
misc_feature	fragment_chain:5"
misc_feature	84715..90326
misc_feature	/note="assembly-fragment:00757"
misc_feature	fragment_chain:5"
misc_feature	90427..92647
misc_feature	/note="assembly-fragment:00677"
misc_feature	fragment_chain:6"
misc_feature	92748..97690
misc_feature	/note="assembly-fragment:00248"
misc_feature	fragment_chain:6"
misc_feature	97791..100799
misc_feature	/note="assembly-fragment:00091"
misc_feature	100900..104202
misc_feature	/note="assembly-fragment:00538"
misc_feature	104303..107157
misc_feature	/note="assembly-fragment:00675"
misc_feature	107258..110300

US-09-830-244B-1 (1-95) x AL450424 (1-146805)

RESULT 4	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	HUMSTRNA	542 bp	mRNA	linear			
	Human statherin mRNA, complete cds.						
	M18371						
	M18371.1	GI:338610					
	statherin.						
	Human female submandibular gland, cDNA, clone pBRHSMST98.2						
	Homo sapiens						

REFERENCE	1 (bases 1 to 542)
AUTHORS	Dickinson, D. P., Riddell, A. L. and Levine, M. J.
TITLE	Human submandibular gland stathmin and basic histidine-rich peptide are encoded by highly abundant mRNA's derived from a common ancestral sequence
JOURNAL	Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)
MEDLINE	86106506
PUBMED	342601
COMMENT	Draft entry and computer readable copy of sequence [1] kindly provided by D.P.Dickinson 21-MAR-1988.
FEATURES	Location/Qualifiers

```

source
Location/Qualifiers
1. .542
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="q11-q13"
1. .542
gene

```

Alignment Scores:	
Pred. No.:	3,356-11
Score:	178.50
Percent Similarity:	63.77%
Best Local Similarity:	53.62%
Query Match:	34,508
DB:	9
Length:	542
Matches:	37
Conservative:	7
Mismatches:	8
Indels:	17
Gaps:	2

OS-09-830-244B-1 (1-95) x HUMSTRNA (1-542)

RESULT	5
HUMSTATHA	
LOCUS	552 bp
DEFINITION	Human statherin mRNA, complete cds.
ACCESSION	M18078
VERSION	M18078.1 GI:338507
KEYWORDS	regulatory protein; statherin.
SOURCE	Human parotid gland, cDNA to mRNA, clone H772B.
ORGANISM	Homo sapiens
PRI	13-JAN-1995

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
1 (bases 1 to 552)	Sabatini, L.M., Carlock, L.R., Johnson, G.W. and Azen, E.A.	CDNA cloning and chromosomal localization (4q11-13) of a gene for statherin, a regulator of calcium in saliva	Am. J. Hum. Genet.	41 (6), 1048-1060 (1987)	88074310	
					3502720	
						Draft entry and computer-readable sequence [1] kindly submitted by L.Sabatini 19-JAN-1988.
						Location/Qualifiers
						1..552
						source

```

1.052
/organism="Homo sapiens
/db_xref="taxon:9606"
/map="4q11-q13"
1.552
/gene="STAT1"
73.261
CDS

```

/gene="STAT1"
/note="statherin protein"
/codon_start=1
/protein_id="AAA60594.1"
/db_xref="GI:338508"
/db_xref="GDB:G00-120-391"
/translation="MKFLVFAPILALVSMIGADSSSEKFLRRIGRGYGYPPVP
NPPLPPQYQFOYQOYTP"
73.129
sig_peptide
/gene="STAT1"
/note="statherin signal peptide"

BASE COUNT 172 a 122 c 78 g 180 t
ORIGIN 278 bp upstream of psi site; chromosome 4q11-q13.

Alignment Scores:

Pred. No.: 3.42e-11 Length: 552
Score: 178.50 Matches: 37
Percent Similarity: 63.77% Conservative: 7
Best Local Similarity: 53.62% Mismatches: 8
Query Match: 34.20% Indels: 17
DB: 9 Gaps: 2

US-09-830-244B-1 (1-95) x HUMSTAT1A (1-552)

OY 27 LeuValSerLeuLysLysLysProPheHisLeuTyrcysValIleTyrlleProLeu 46
DB 128 CTGATTCATCTGAAGAGAAAT-----TTTCCGTA----- 157
OY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
DB 158 -----/-----GAATTGGAGATTCCGTTATGGGTATGGCCCTTATCAGC 196
OY 67 GluPheGlnAsnAsnHisTyrrThHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
DB 197 CAGTTCGAGAACCAACCATATACCCACACCATACCAACCAATACCAATATATACCT 256
OY 87 PheAsnIleIleSerAsnCysArgThr 95
DB 257 TTTAATATCATCACTAAGTCAAGACA 283

RESULT 6 1584 bp mRNA linear PRI 15-JUL-2002
AK092678 Homo sapiens cDNA FL35359 fts, clone SALGL1000107, highly similar
LOCUS AK092678
DEFINITION Homo sapiens cDNA FL35359 fts, clone SALGL1000107, highly similar
TO CH-TCG PROTEIN.

ACCESSION AK092678.1 GI:21751327
VERSION AK092678.1
KEYWORDS oligo capping; fts (full insert sequence)
SOURCE Homo sapiens salivary gland cDNA to mRNA, clone lib: SALGL1
clone: SALGL1000107.
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
Arita, M., Musashiro, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
Suzuki, Y., Sugano, S., Nagahori, K., Masuno, Y., Nagai, K., and
Isogai, T.
TITLE NED0 human cDNA sequencing project
JOURNAL unpublished
REFERENCE 2 (bases 1 to 1584)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FRI Project(HRI Team); 2-6-7

COMMENT (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NED0 human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology (RAB); cDNA library

Construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing; HRI and
RAB; annotation: HRI and RAB.

FEATURES

source
1.1584
Location/Qualifiers

CDS

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="SALGL1000107"
/tissue="salivary gland"
/clone_lib="SALGL1"
/note="Cloning vector: pME18SF13"
89..997
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC03943.1"
/db_xref="GI:21751328"
/translation="MGDDSEWLKLPVOKCEKRLKRLKSGEALKIPKIDKSP
EMSKFLGIKKEVDSNANVYOLKGLAALVYENAHVAGTTGVSQVSKFNOR
AKAEICETICMTETIEKRAVEELKGLDNKPKIIVACIETTLKALSEFSKII
ILKPIIKVPLKESREKAVDEAKLIVELIYRWIRDALPPIQINISVOLKLEEM
VLPFSAPRPLRFSQOELEAKLEDOOSAGGADGGDDGDEVPQIDAVELLEAVEI
LSKLPEKFMRLRKQKGRKRPMSL"
BASE COUNT 526 a 298 c 331 g 429 t
ORIGIN

Alignment Scores:

Pred. No.: 1.02e-10 Length: 1584
Score: 178.50 Matches: 37
Percent Similarity: 63.77% Conservative: 7
Best Local Similarity: 53.62% Mismatches: 8
Query Match: 34.20% Indels: 17
DB: 9 Gaps: 2

US-09-830-244B-1 (1-95) x AK092678 (1-1584)

OY 27 LeuValSerLeuLysLysLysProPheHisLeuTyrcysValIleTyrlleProLeu 46
DB 1152 CTGATTCATCTGAAGAGAAAT-----TTTCCGTA----- 1181
OY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
DB 1182 -----GAATTGGAGATTCCGTTATGGGTATGGCCCTTATCAGC 1220
OY 67 GluPheGlnAsnAsnHisTyrrThHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
DB 1221 CAGTTCGAGAACCAACCATATACCCACACCATACCAACCAATATATACCT 1280
OY 87 PheAsnIleIleSerAsnCysArgThr 95
DB 1281 TTTAATATCATCACTAAGTCAAGACA 1307

RESULT 7 4723 bp DNA linear PRI 13-JAN-1995
HUMSTAT2
LOCUS HUMSTAT2
DEFINITION Human salivary statherin gene, exons 2-6.
ACCESSION M32639.1 GI:338504
VERSION M32639.1
KEYWORDS statherin.
SOURCE Human (individuals #563, #8136, and J.F.) fibroblast, cell line
#563, DNA, clones 1-3.
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Sabatin, L.M., He, Y.Z., and Azen, E.A.
JOURNAL Structure and sequence determination of the gene encoding human
salivary statherin
MEDLINE Gene 89 (2), 245-251 (1990)
PUBMED 90323623
2373369

COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by L.M.Sabatini, 03-JAN-1990.

FEATURES Location/Qualifiers

source 1. 4723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="4q11-q13"
/contig="31077.1:1601..1931.1..352"
/gene="STATH"
/note="STATH mRNA and introns"
1. 286
/gene="STATH"
/note="STATH intron A"
/contig="31077.1:1601..1931.1..352"
/note="stath intron precursor"
/codon_start=1
/protein_id="AA060593.1"
/db_xref="GI:338506"
/translation="MKLVFAFIALMYSMTGADSSSEKFLRRIGRGYGYQPYVP
EQLYRQPTQPYQYTP"
/contig="31077.1:1601..1931.1..352"
/note="stath intron signal peptide"
/contig="31077.1:1601..1931.1..352"
/product="stath intron"
302..352
/gene="STATH"
/note="stath intron precursor, (first translated exon);
500-120-391"
/number=2
353..1531
/note="STATH intron B"
1532..1552
/number=3
1533..1644
/note="STATH intron C"
1645..1674
/number=4
1675..2727
/note="STATH intron D"
2728..2814
/note="stath intron precursor"
/number=5
2848..4051
/note="STATH intron E"
4294..4299
polyA_signal 1570 a 722 g 1634 t
BASE COUNT About 1.8 kb after segment 1.
ORIGIN

Alignment Scores:
Pred. No.: 3.58e-10 Length: 4723
Score: 178.00 Matches: 34
Percent Similarity: 77.08% Conservative: 3
Best Local Similarity: 70.83% Mismatches: 5
Query Match: 34.10% Indels: 6
DB: 9 Gaps: 1

US-09-830-244b-1 (1-95) x HUMSTATH2 (1-4723)

Oy 54 Pheleuaplelelaph-----lleProlysSerleuileSerGln 67
Db 2693 TTTCGCAATTTCTCTCTCTGCTATACAGATATGGGTATGCGCTTATCAGCAG 2752

Oy 68 PheGlnAsnAsnHsTYrThFhAsnHsThrsnHsAsnThrsnHsAsnHsAsnHsAsnHs 87
Db 2753 TTTCAGAAACACACTATACCAACATACCAACATACCAACATATATATATATATAT 2812

Oy 88 AsnIleIleSerAsnGysArgThr 95
Db 2813 AATATCATCATGTAATGTCAGAGACA 2836

RESULT 8
AC063956

LOCUS AC063956 141568 bp DNA linear PRI 25-AUG-2000
DEFINITION Homo sapiens 4 BAC Rpl1-529K3 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
AC063956
VERSION AC063956.7 GI:9910030
KEYWORDS HTS.
SOURCE
ORGANISM Homo sapiens.
Homo sapiens.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 141568)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Aye,J.R., Banks,T., Barbier,J.,
Benton,U., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burckel,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferriguto,D., Flaggs,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunatane,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
Hollaway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulik,S.,
Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landy,N.,
Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Liu,C.,
Liu,J., Liu,M., Louised,H., Lozada,R.J., Lu,X., Lucier,A.,
Lucier,R., Luna,R., Ma,J., Maneshwar,M., Mapua,P., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,
Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T.,
Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenko,S., Ogun,K., Okunodu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabot,P., Tameris,A., Tameris,K., Tang,H.,
Tansey,J., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
Naylor,S.L. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 141568)
Worley,K.C.
Direct Submission
Submitted (22-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 141568)
Worley,K.C.
Direct Submission
Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 25, 2000 this sequence version replaced g1:9795448.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

SSIs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats were identified using RepeatMasker A. Smit and P. Green (unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: this sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics

Concat length:	141568
Phrap values in estimate:	140895
Average error rate (BCM-phrap estimate):	1.4137e-05
Fraction of Phrap values less than 40 :	0.00715426
Number of consensus changing edits:	10
Number of N's in consensus :	0

[illegible]

```
----- Distribution of Quality < 40 Bases -----
```

#	bases	Phrap Value Range
10001		
9001		
8001		
7001		
6001		
5001		
4001		
3001		
2001		
1001		
01		

```

Version: 1.01  qxfc.
FEATURES      Location/Qualifiers
source        1. .141568

```

Version: 1.01 qxf0.

repeat_region	969..1013	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="4" /clone="Rp11-529K3"
repeat_region	1685..1904	/rpt_family="AT_rich" /rpt_family="L2"
repeat_region	2500..2548	/rpt_family="AT_rich"
repeat_region	2598..2647	/rpt_family="AT_rich"
repeat_region	3488..5870	/rpt_family="LTPA" complement(5909..6044) /rpt_family="L2"
repeat_region	6890..6913	/rpt_family="AT_rich"
repeat_region	7539..7560	/rpt_family="TA)n"
repeat_region	complement(8596..8810)	/rpt_family="MIR"
repeat_region	8845..8865	/rpt_family="AT_rich"
repeat_region	10335..10727	/rpt_family="LMA9"
repeat_region	12829..13116	/rpt_family="AluY"
repeat_region	13141..13282	/rpt_family="LTPA5"
repeat_region	complement(14166..14219)	/rpt_family="MER20"
repeat_region	14224..14291	/rpt_family="T89ge1"
repeat_region	complement(14336..17747)	/rpt_family="LTPA5"
repeat_region	complement(17748..18164)	/rpt_family="MSYA"
repeat_region	complement(18165..18621)	/rpt_family="LTPA5"
repeat_region	18622..18644	/rpt_family="(TTTG)n"
repeat_region	complement(18645..18922)	/rpt_family="LTPA5"
repeat_region	18923..18944	/rpt_family="(TG)n"
repeat_region	complement(18945..21150)	/rpt_family="LTPA5"
repeat_region	complement(21216..21381)	

Alignment Scores:

pred. NO.:	1.21e-08	length:	141566
Score:	178.00	Matches:	34
Percent Similarity:	77.08%	Conservative:	3
Best Local Similarity:	70.83%	Mismatches:	5
Query Match:	34.10%	Indels:	1
DB:	9	Gaps:	6

US-09-830-244B-1 (1-95) x AC063956 (1-141568)

```

27      34 FHELEUASPRILEALAPHE-----ILEPROLYSSERLEULESERGLN 67
      |||||      ::|||      ::|||

```

On 60 Dec 2011, 11:00 AM, 43016

Db 43017 TTCCAGAACACCACCTATACCCACAACCATACCAACACATACCAACAATTTACCTTTT 43076

OY 88 Asn1le1leSerAsnCysArgThr 95
 DB 43077 AATATCATCACTACTGACAGACA 43100
 RESULT 9
 AC024676 161549 bp DNA linear HTG 16-MAR-2000
 LOCUS
 DEFINITION Homo sapiens chromosome 4 clone RP11-751H12 map 4, WORKING DRAFT
 AC024676
 AC024676.2 GI:7249372
 HTG: HTGS_PHASE1: HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 161549)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 4, clone RP11-751H12
 Unpublished
 2 (bases 1 to 161549)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domano, M., Doyle, M., Ferrelle, P., FitzHugh, M., Gage, D.,
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McKern, A., McKernan, K., McPherson, R.,
 Meltzer, J., Meneses, L., Milnova, T., Miranda, C., Mlenka, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Tamas, J.,
 Testave, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced gi:7139892.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 1060: contig of 1060 bp in length
 * 1061 1160: gap of 100 bp
 * 1161 2376: contig of 1216 bp in length
 * 2377 2476: gap of 100 bp
 * 2477 3989: contig of 1513 bp in length
 * 3990 4089: gap of 100 bp
 * 4090 5305: contig of 1216 bp in length
 * 5306 5405: gap of 100 bp
 * 5406 7418: contig of 2013 bp in length
 * 7419 7518: gap of 100 bp
 * 7519 9669: contig of 2151 bp in length
 * 9670 9769: gap of 100 bp
 * 9770 11064: contig of 1295 bp in length
 * 11065 11164: gap of 100 bp
 * 11165 14241: contig of 3077 bp in length
 * 14242 14341: gap of 100 bp
 * 14342 17832: contig of 3491 bp in length
 * 17833 17932: gap of 100 bp
 * 17933 22345: contig of 4413 bp in length
 * 22346 22445: gap of 100 bp
 * 22446 27054: contig of 4609 bp in length
 * 27055 27154: gap of 100 bp
 * 27155 31430: contig of 4276 bp in length
 * 31431 31530: gap of 100 bp
 * 31531 37872: contig of 6342 bp in length
 * 37873 37972: gap of 100 bp
 * 37973 44394: contig of 6422 bp in length
 * 44395 44494: gap of 100 bp
 * 44495 51121: contig of 6627 bp in length
 * 51122 51221: gap of 100 bp
 * 51222 58058: contig of 6837 bp in length
 * 58059 58158: gap of 100 bp
 * 58159 67135: contig of 8977 bp in length
 * 67136 67235: gap of 100 bp
 * 67236 75851: contig of 8616 bp in length
 * 75852 75951: gap of 100 bp
 * 75952 84245: contig of 8294 bp in length
 * 84246 84345: gap of 100 bp
 * 84346 95422: contig of 11077 bp in length
 * 95423 95522: gap of 100 bp
 * 95523 108795: contig of 13273 bp in length
 * 108796 108895: gap of 100 bp
 * 108896 135384: contig of 26489 bp in length
 * 135385 135484: gap of 100 bp
 * 135485 161549: contig of 26065 bp in length.
 FEATURES
 source
 1..161549
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-751H12"
 /clone_id="RP11-751H12 Human Male BAC"
 1..1060
 /note="assembly-fragment"
 1161..2376
 /note="assembly-fragment"
 2477..3989
 /note="assembly-fragment"
 4090..5305
 /note="assembly-fragment"
 5406..7418
 /note="assembly-fragment"
 7519..9669
 /note="assembly-fragment"
 9770..11064
 /note="assembly-fragment"
 clone_end:T7
 vector_side:left

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center Project name: L7432
 Center Clone name: 751_H_12
 Summary Statistics
 Sequencing vector: M13: M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 148549 bases at least Q40
 Consensus quality: 14910 bases at least Q30
 Consensus quality: 157453 bases at least Q20
 Insert size: 157000; agarose-fp
 Insert size: 159349; sum-of-contigs
 Quality coverage: 3.8 in Q20 bases; agarose-fp
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs
 NOTE: This is a 'working draft' sequence. It currently
 consists of 23 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as

```

misc_feature      11165..14241
                   /note="assembly-fragment"
misc_feature      14342..17832
                   /note="assembly-fragment"
misc_feature      17933..22345
                   /note="assembly-fragment"
misc_feature      22446..27054
                   /note="assembly-fragment"
misc_feature      27155..31430
                   /note="assembly-fragment"
misc_feature      31531..37872
                   /note="assembly-fragment"
misc_feature      37973..44394
                   /note="assembly-fragment"
misc_feature      44495..51121
                   /note="assembly-fragment"
misc_feature      51222..58058
                   /note="assembly-fragment"
misc_feature      58159..67135
                   /note="assembly-fragment"
misc_feature      67236..75851
                   /note="assembly-fragment"
misc_feature      75952..84245
                   /note="assembly-fragment"
misc_feature      84346..95422
                   /note="assembly-fragment"
misc_feature      95523..108795
                   /note="assembly-fragment"
misc_feature      108896..135384
                   /note="assembly-fragment"
misc_feature      135485..161549
                   /note="assembly-fragment"
BASE COUNT      53276 a 27644 c 27367 g 51058 t 2204 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.38e-08      Length:      161549
Score:          178.00      Matches:      34
Percent Similarity: 77.08%      Conservative: 3
Best Local Similarity: 70.83%      Mismatches: 5
Query Match:     34.10%      Indels:      6
                        Gaps:      1

```

US-09-830-244B-1 (1-95) x AC024676 (1-161549)

```

QY      54 Phlebaspi1le1aphe-----1leProlyserleu1leSergin 67
        |||||      |||||      |||||      |||||      |||||
DB 125221 TTCTGCAATTTCTTCTCTGTTGTTATACAGTATGCGCTTATACAGCCAG 125280
        |||||      |||||      |||||      |||||      |||||
QY      68 Phcglanasn1st1ythr1h1saen1st1h1asn1st1h1asn1le1ar1gpe 87
        |||||      |||||      |||||      |||||      |||||
DB 125281 TTCCAGAACACACATATACCCACACATACACACATACACATACACATACCTTT 125340
        |||||      |||||      |||||      |||||      |||||
QY      88 Asn1le1leSern1ncysar1gthr 95
        |||||      |||||      |||||      |||||      |||||
DB 125341 AATATCATCATGTAATGACAGACA 125364

```

```

RESULT 10
AC021462/c
LOCUS      AC021462      153023 bp      DNA      linear      HTG 03-APR-2000
DEFINITION Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
pieces:
AC021462
VERSION      AC021462.3      GI:7387343
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 153023)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens, clone RP11-276C1

Unpublished
2 (bases 1 to 153023)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckel,R., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearlano,K., Dewar,K., Domino,M., Doyle,W., Fenebor,D.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,D.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lechoczky,J., Levine,R., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McCurt,A., McKernan,K.,
McPheters,R., Meldrum,J., Menus,L., Morrill,D., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Ollivar,T.M., Peterson,K.,
Pierre,N., Pisanil,C., Pollara,V., Raymond,C., Riley,R., Roehman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

TITLE JOURNAL

COMMENT

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7230200.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 15201

Center clone name: 276C_1

Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141815 bases at least Q40
Consensus quality: 147368 bases at least Q30
Consensus quality: 149481 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 151023; sum-of-ctrls
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
consists of 21 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
1016 1115: gap of 100 bp in length
1116 2276: contig of 1161 bp in length
2277 2376: gap of 100 bp in length
2377 4009: contig of 1633 bp in length
4010 4109: gap of 100 bp in length
4110 5737: contig of 1678 bp in length
5738 5837: gap of 100 bp in length
5838 6939: contig of 1102 bp in length
6940 7039: gap of 100 bp in length
7040 9504: contig of 2465 bp in length
9505 9604: gap of 100 bp in length
9605 11734: contig of 2130 bp in length
11735 11834: gap of 100 bp in length
11835 14111: contig of 2277 bp in length
14112 14211: gap of 100 bp in length
14212 16973: contig of 2762 bp in length
16974 17073: gap of 100 bp in length

```

* 17074 19588: contig of 2515 bp in length
* 19589 19688: gap of 100 bp
* 19689 22275: contig of 2587 bp in length
* 22276 22375: gap of 100 bp
* 22376 25823: contig of 3448 bp in length
* 25824 25923: gap of 100 bp
* 25924 31307: contig of 5384 bp in length
* 31308 31407: gap of 100 bp
* 31408 34863: contig of 3456 bp in length
* 34864 34963: gap of 100 bp
* 34964 41382: contig of 6419 bp in length
* 41383 41482: gap of 100 bp
* 41483 49024: contig of 7542 bp in length
* 49025 49124: gap of 100 bp
* 49125 58672: contig of 9548 bp in length
* 58673 58772: gap of 100 bp
* 58773 69622: contig of 10850 bp in length
* 69623 69722: gap of 100 bp
* 69723 88191: contig of 18465 bp in length
* 88192 88291: gap of 100 bp
* 88292 107084: contig of 18793 bp in length
* 107085 107184: gap of 100 bp
* 107185 153023: contig of 45839 bp in length.

```

FEATURES

source

```

misc_feature /organism="Homo sapiens"
misc_feature /db_xref="taxon:9606"
misc_feature /clone="RP11-276C1"
misc_feature /clone_id="RP11-276C1"
misc_feature 1. 1015
misc_feature /note="assembly-fragment"
misc_feature 1116. 2276
misc_feature /note="assembly-fragment"
misc_feature 2377. 4009
misc_feature /note="assembly-fragment"
misc_feature 4110. 5737
misc_feature /note="assembly-fragment"
misc_feature 5838. 6939
misc_feature /note="assembly-fragment"
misc_feature 7040. 9504
misc_feature /note="assembly-fragment"
misc_feature 9605. 11734
misc_feature /note="assembly-fragment"
misc_feature 11835. 14111
misc_feature /note="assembly-fragment"
misc_feature 14212. 16973
misc_feature /note="assembly-fragment"
misc_feature 17074. 19588
misc_feature /note="assembly-fragment"
misc_feature 19689. 22275
misc_feature /note="assembly-fragment"
misc_feature 22376. 25823
misc_feature /note="assembly-fragment"
misc_feature 25924. 31307
misc_feature /note="assembly-fragment"
misc_feature 31408. 34863
misc_feature /note="assembly-fragment"
misc_feature 34964. 41382
misc_feature /note="assembly-fragment"
misc_feature 41483. 49024
misc_feature /note="assembly-fragment"
misc_feature 49125. 58672
misc_feature /note="assembly-fragment"
misc_feature 58773. 69622
misc_feature /note="assembly-fragment"
misc_feature 69723. 88191
misc_feature /note="assembly-fragment"
misc_feature 88292. 107084
misc_feature /note="assembly-fragment"
misc_feature 107185. 153023
misc_feature /note="assembly-fragment"
misc_feature clone_end:17
misc_feature vector_side:left"

```

```

BASE COUNT 41494 a 35192 c 34573 g 39756 t 2008 others
ORIGIN
Alignment Scores:
Pred. No.: 9.39e-05
Score: 143.50
Percent Similarity: 53.33%
Best Local Similarity: 41.33%
Query Match: 27.49%
DB: 2
Gaps: 1
US-09-830-244B-1 (1-95) x AC021462 (1-153023)
QY 1 MetTTPheHISLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
|||||
Db 13988 ATGTGTTTCATCAACAGTGGAGAAAGCAGCATTTAAATACCTTTTGGAGACTGAA 13929
|||||
QY 21 LeuSerAsnAsnLysThr
|||||
Db 13928 TTGAGTAATATAAATACTTCACTTTCGCTAATATAATAATAATAATAACAA 13869
|||||
QY 27 -----LeuValSerLeuLysLysLysProPheHisLeuTyr 39
|||||
Db 13868 CAACTTATGATGTGGCGACGCTCAGTACAGAGAAAGAGAGCATTTTCTGCATTC 13809
|||||
QY 40 -CysValIleTyrIleProLeuValProLysLeuIleLeu 53
|||||
Db 13808 TTGCCTAGTTTCCTTATAGCACCCATACATTAATAGCTCTG 13766
|||||
RESULT 11
AX281311
LOCUS AX281311
DEFINITION Sequence 53 from Patent WO0177164.
ACCESSION AX281311
VERSION AX281311.1 GI:16608566
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with apoptosis
JOURNAL Patent: WO 0177164-A 53 18-OCT-2001;
Epidemiomics AG (DE)
FEATURES
LOCATION/Qualifiers
SOURCE 1. 9881
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2606 a 196 c 2308 g 4771 t
ORIGIN
Alignment Scores:
Pred. No.: 0.00574
Score: 116.50
Percent Similarity: 51.32%
Best Local Similarity: 36.84%
Query Match: 22.32%
DB: 6
Gaps: 1
US-09-830-244B-1 (1-95) x AX281311 (1-9881)
QY 1 MetTTPheHISLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
|||||
Db 5782 ATGTGTTTATTAAGTGGAGAAAGTACATTAATTAATTTTGGAGACTGAA 5841
|||||
QY 21 LeuSerAsnAsnLysThrLeuValSerLeu
|||||
Db 5842 TTGAGTAATATAAATACTTTCGCTAATATAATAATAATAATAATAATAA 5901
|||||
QY 31 -----LysLysLysLysProPheHisLeu-Ty 39
|||||
Db 5902 TAAATTATGAATGCTGTTAGTTTATTAAGTACGAGAAAGAGAGCATTTTGTGATTT 5961
|||||

```

TITLE	Genomic organization of the <i>hmd</i> gene
Direct Submission	Demchenko, D., Gilmer, A. and Zody, M.
Submitted (20-JUN-2002)	Whitehead Institute/MIT Center for Genome Research

1-148626)

05-03-830-244B-1 (1-95) x AC115931 (1-146828)

QY 1 MetTPRPhHisILysValGlyArgLysGlnHisPheLysVal---ThirperrGlnThr 19
||||| ||| ||| ::|||::: ||| |||

Db 131844 ATGTGATTCACACGCTTTGGGCAAAATGAACCTCAGGACTTAAGCAGTTAT-GAAGCA 131902
 Oy 20 AspleuseaAsnAsnlyThleuVal---SerleuIys----- 31
 Db 131903 GACGTGGATCTATATTAAGATTATTAATCTATCTTATTAAGTAAGAAACAGCATACCT 131962
 Oy 32 -----LysIysIysProPheHs-----Leu 38
 Db 131963 GAGACATGGGATGAGCGCTTCCTCAAGAGACACCTATCTCAGTATGTTTGTGTTTG 132022
 Oy 39 TyrGysValIleTyrIleProIeuValProIysLeuIleIleIleuPheIeuAspIleAla 58
 Db 132023 TTTGTTGTTTTCACACATTAACATATACAGATTTGTTGAGCGCTTGTACTATATTTT 132082
 Oy 59 PheIleProIysSerIleIleSerGlnPheGlnAsnAsnHs-----TyrThrHis 75
 Db 132083 -----AAAGGCTTCTTAAGACACATAAATGAGACACACACACACACACACACACAC 132133
 Oy 76 AsnHisThrAsnHisAsnThrAsnAsnHis 85
 Db 132134 ACACACACACACACACACACACACACAGTGT 132163

RESULT 13
 AC100542 55219 bp DNA linear HTG 22-NOV-2001
 LOCUS AC100542
 DEFINITION Mus musculus clone RP23-152620, LOW-PASS SEQUENCE SAMPLING.
 AC100542
 VERSION AC100542.1 GI:17047908
 KEYWORDS HTG: HTGS_PHASD.
 SOURCE Mus musculus.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 55219)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Mus musculus, clone RP23-152620
 REFERENCE 2 (bases 1 to 55219)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choquel, Y., Collange, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeBellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farcot, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., Lacombe, K.,
 Iamazares, R., Landers, T., Lehotzky, V., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
 Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollard, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Semman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tothman, K., Travers, M., Travis, N., Triggillo, J., Vassiliou, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
 COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WITR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L15664

Center clone name: 152_G_20

* NOTE: This record contains 69 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 729: contig of 729 bp in length
 * 730 829: gap of 100 bp
 * 830 1532: contig of 703 bp in length
 * 1533 1632: gap of 100 bp
 * 1633 2340: contig of 708 bp in length
 * 2341 2440: gap of 100 bp
 * 2441 3151: contig of 711 bp in length
 * 3152 3251: gap of 100 bp
 * 3252 3940: contig of 689 bp in length
 * 3941 4040: gap of 100 bp
 * 4041 4731: contig of 691 bp in length
 * 4732 4831: gap of 100 bp
 * 4832 5517: contig of 686 bp in length
 * 5518 5617: gap of 100 bp
 * 5618 6321: contig of 704 bp in length
 * 6322 6421: gap of 100 bp
 * 6422 7135: contig of 714 bp in length
 * 7136 7235: gap of 100 bp
 * 7236 7967: contig of 732 bp in length
 * 7968 8067: gap of 100 bp
 * 8068 8795: contig of 728 bp in length
 * 8796 8895: gap of 100 bp
 * 8896 9582: contig of 687 bp in length
 * 9583 9682: gap of 100 bp
 * 9683 10388: contig of 706 bp in length
 * 10389 10488: gap of 100 bp
 * 10489 11188: contig of 700 bp in length
 * 11189 11288: gap of 100 bp
 * 11289 11975: contig of 687 bp in length
 * 11976 12075: gap of 100 bp
 * 12076 12765: contig of 690 bp in length
 * 12766 12865: gap of 100 bp
 * 12866 13566: contig of 701 bp in length
 * 13567 13666: gap of 100 bp
 * 13667 14362: contig of 696 bp in length
 * 14363 14462: gap of 100 bp
 * 14463 15163: contig of 701 bp in length
 * 15164 15263: gap of 100 bp
 * 15264 15958: contig of 695 bp in length
 * 15959 16058: gap of 100 bp
 * 16059 16758: contig of 700 bp in length
 * 16759 16858: gap of 100 bp
 * 16859 17561: contig of 703 bp in length
 * 17562 17661: gap of 100 bp
 * 17662 18379: contig of 718 bp in length
 * 18380 18479: gap of 100 bp
 * 18480 19153: contig of 674 bp in length
 * 19154 19253: gap of 100 bp
 * 19254 19950: contig of 697 bp in length
 * 19951 20050: gap of 100 bp
 * 20051 20739: contig of 689 bp in length
 * 20740 20839: gap of 100 bp
 * 20840 21538: contig of 693 bp in length
 * 21539 21638: gap of 100 bp
 * 21639 22336: contig of 698 bp in length
 * 22337 22436: gap of 100 bp
 * 22437 23145: contig of 709 bp in length
 * 23146 23245: gap of 100 bp
 * 23246 23957: contig of 712 bp in length
 * 23958 24057: gap of 100 bp

24058	24788:	contlig of 731 bp	in length
24789	24888:	gap of 100 bp	
24889	25597:	contlig of 709 bp	in length
25598	25697:	gap of 100 bp	
25698	26400:	contlig of 703 bp	in length
26401	26500:	gap of 100 bp	
26501	27188:	contlig of 688 bp	in length
27189	27288:	gap of 100 bp	
27289	27990:	contlig of 702 bp	in length
27991	28090:	gap of 100 bp	
28091	28774:	contlig of 684 bp	in length
28775	28874:	gap of 100 bp	
28875	29591:	contlig of 717 bp	in length
29592	29691:	gap of 100 bp	
29692	30427:	contlig of 736 bp	in length
30428	30527:	gap of 100 bp	
30528	31223:	contlig of 698 bp	in length
31226	31325:	gap of 100 bp	
31326	32039:	contlig of 714 bp	in length
32040	32139:	gap of 100 bp	
32140	32841:	contlig of 702 bp	in length
32842	32941:	gap of 100 bp	
32942	33624:	contlig of 683 bp	in length
33625	33724:	gap of 100 bp	
33725	34415:	contlig of 691 bp	in length
34416	34515:	gap of 100 bp	
34516	35221:	contlig of 706 bp	in length
35222	35321:	gap of 100 bp	
35322	36031:	contlig of 710 bp	in length
36032	36131:	gap of 100 bp	
36132	36885:	contlig of 727 bp	in length
36885	36958:	gap of 100 bp	
36959	37684:	contlig of 726 bp	in length
37685	37784:	gap of 100 bp	
37785	38481:	contlig of 697 bp	in length
38482	38581:	gap of 100 bp	
38582	39282:	contlig of 701 bp	in length
39283	39382:	gap of 100 bp	
39383	40077:	contlig of 665 bp	in length
40078	40177:	gap of 100 bp	
40178	40845:	contlig of 668 bp	in length
40846	40945:	gap of 100 bp	
40946	41633:	contlig of 678 bp	in length
41624	41723:	gap of 100 bp	
41724	42433:	contlig of 710 bp	in length
42434	42533:	gap of 100 bp	
42534	43244:	contlig of 711 bp	in length
43245	43344:	gap of 100 bp	
43345	44050:	contlig of 706 bp	in length
44051	44150:	gap of 100 bp	
44151	44847:	contlig of 697 bp	in length
44848	44947:	gap of 100 bp	
44948	45656:	contlig of 709 bp	in length
45657	45756:	gap of 100 bp	
45757	46470:	contlig of 714 bp	in length
46471	46570:	gap of 100 bp	
46571	47262:	contlig of 692 bp	in length
47263	47362:	gap of 100 bp	
47363	48041:	contlig of 679 bp	in length
48042	48141:	gap of 100 bp	
48142	48843:	contlig of 702 bp	in length
48844	48943:	gap of 100 bp	
48944	49638:	contlig of 695 bp	in length
49639	49738:	gap of 100 bp	
49739	50416:	contlig of 678 bp	in length
50417	50516:	gap of 100 bp	
50517	51244:	contlig of 728 bp	in length
51245	51344:	gap of 100 bp	
51345	52051:	contlig of 707 bp	in length
52052	52151:	gap of 100 bp	
52152	52843:	contlig of 692 bp	in length
52844	52943:	gap of 100 bp	
52944	53621:	contlig of 678 bp	in length

100

```

misc_feature      1..194709
                  /note="assembly_fragment:00367"
misc_feature      194810..213680
                  /note="assembly_fragment:03782"
misc_feature      213781..215984
                  /note="assembly_fragment:03818"
misc_feature      216085..218926
                  /note="assembly_fragment:03999"
misc_feature      219027..221365
                  /note="assembly_fragment:04053"
BASE COUNT      64278 a 42370 c 42725 g 71592 t 400 others
ORIGIN

```

Alignment Scores:

```

Pred. No.:      5.25      Length:      221365
Score:          102.50     Matches:      33
Percent Similarity:      44.44%     Conservative: 15
Best Local Similarity:    30.56%     Mismatches:   35
Query Match:      19.64%     Indels:      25
DB:              2        Gaps:      5

```

US-09-830-244b-1 (1-95) x AL662923 (1-221365)

```

OY      2  TrrPheHISlySValGlyArgLysGlnHISpHeLyValThr-----PheTrrGlu 18
          |||      ::      |||      |||      |||      |||      |||
Db 164436 TGGTGCCCAAGAGAGGAAATCCACCCCTCTGTGAAAAGAACAGATGGGGAAATGGAG 164377
OY      19  ThrAsp-----LeuSerAsnAsnLysThrLeuValSerLeu-----30
          |||      |||      ::      |||      ::      |||      ::
Db 164376 GAAGACTTTTAAAGGAAAGACTTAAGAGAGGAGGAGACTGTGATCATGATGTAAAGTGAA 164317
OY      31  -----LysLysLysLysProPheHISLeuTyrCysValIleTyrIle 44
          |||      |||      |||      |||      |||      |||      |||
Db 164316 TAAATAATAACTATTGGAAGAAAGTAAGAACCAAGCTTTATATTGT-----TATATA 164263
OY      45  ProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeu 64
          |||      |||      |||      |||      |||      |||      |||
Db 164262 GAATTAGTGCCA-----CTACAGATATTTTTCTACACTTTATTGG 164221
OY      65  IleSerGlnPheGlnAsnAsnHISyrrHISAsnHISThrAsnHISAsnThrAsnAsn 84
          |||      |||      ::      |||      |||      |||      |||
Db 164220 ATAGCTGTTTACACACACACACACACACACACACACACACACACACACACACATA 164161
OY      85  IleArgPheAsnIleIleSerAsn 92
          |||      |||      ::      |||      ::      |||      ::
Db 164160 CACACACATAAATTTTAAATTCA 164137

```

Search completed: April 14, 2003, 23:39:49
 Job time : 1553 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2003, 23:05:11 ; Search time 225 Seconds

(Without alignments)
950.844 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522
Sequence: 1 MWFHKVGRKQHFXTFWETD.....NHTNHTNNTNIRFNISNCR 95

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV-xip
-O/cgn2.1/USPTO.spool/US09830244/runat_10042003_090015_27856/app_query.fasta_1.263
-DB-N_Geneseq_101002 -OPT-fastlap -SUFFIX-rng -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -DOCLIG=200 -THR-SCORE-pct -THR-MAX=100 -THR-MIN=0 -ALIGN=15
-USRR-LOCAL -OUTFMT-plo -NORM-ext -HEAPSIZE=500 -MNTLEN=0 -MAXLEN=2000000000
-NO_XLPHY -NO_MMAPP -LARGEJOURN -NEG_SCORES=0 -NAIT -LONGLOG -DEV-TIMEOUT=120
-MARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	522	100.0	1331	21	AAA48963	Human lysine-rich
2	200	38.3	221	21	AAA42355	Human secreted exp
3	178.5	34.2	552	21	AAA48964	Human statherin DN
4	178.5	34.2	2121	21	AAC76665	Human ORFX ORF220
5	136	26.1	390	22	AA191635	Human polynucleotl
6	121.5	23.3	516	20	AAV89555	EST clone CP289.
7	117	22.4	126	16	AA124320	Human gene signatu
8	116.5	22.3	9881	24	AB154353	Chemically treated
9	113	21.6	60	24	ABN43273	Human spliced tran
10	101	19.3	6070	24	AB192119	Chemically treated
11	101	19.3	6070	24	AB192119	Human polynucleotl
12	101	19.3	6070	24	AB132241	Human immune syste
13	92	17.6	9021	22	AA546326	Tumour suppressor
14	91.5	17.5	7038	22	AA545496	Chemically pretrea
15	91.5	17.5	7038	24	AB134097	Human immune syste
16	91.5	17.5	7038	24	ABK28426	DNA transcription
17	91.5	17.5	9881	24	ABK28426	Chemically treated
18	89.5	17.1	8246	24	AB154354	Human gene regulat
19	87.5	16.8	32205	22	ABJ32202	Chemically treated
20	87.5	16.8	32205	22	ABJ32202	Human immune syste
21	87.5	16.8	32205	22	ABJ32202	Human ovarian and
22	86	16.5	6754	24	AA561305	Human reproductive
23	86	16.5	6754	24	AA561305	Wild-type human CT
24	85	16.3	6151	24	AB132622	Chemically treated
25	83.5	16.0	1136	21	AAC59338	Human gene regulat
26	83.5	16.0	6535	24	ABJ32936	Human immune syste
27	83.5	16.0	12571	21	AA352722	Human secreted pro
28	82.5	15.8	16545	24	ABJ32051	Human immune syste
29	82.5	15.8	1791	23	AB105983	Soybean retroviral
30	82.5	15.8	8889	23	AB105982	Human immune syste
31	82	15.7	8889	24	AB132211	Drosophila melanog
32	81.5	15.6	6031	22	AA546522	Human immune syste
33	81.5	15.6	13376	24	AB132582	Tumour suppressor
34	80	15.3	556	22	AA191773	Human immune syste
35	80	15.3	5666	24	AB133472	Human polynucleotl
36	80	15.3	6059	24	AB133472	Human immune syste
37	80	15.3	6350	24	AB170500	Human immune syste
38	80	15.3	6350	24	AB134629	Chemically treated
39	79.5	15.2	13814	24	AB133193	Human metastasis a
40	79	15.1	8996	22	AA545503	Human immune syste
41	79	15.1	8996	24	ABK28435	Chemically pretrea
42	78.5	15.0	10460	23	AB103520	DNA transcription
43	78.5	15.0	10514	23	AB120932	Drosophila melanog
44	78.5	15.0	17722	23	AB103508	Drosophila melanog
45	78.5	15.0	30037	22	ABA20038	Human nervous syst

ALIGNMENTS

RESULT 1
AAA48963
ID AAA48963 standard; cDNA; 1331 BP.

AAA48963;
06-OCT-2000 (first entry)

Human lysine-rich statherin cDNA from Incyte clone 2820214.

Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
asthma; allergy; diabetes mellitus; fungal; bacterial infection;
cancer; leukemia; adenocarcinoma; melanoma; ss.

OS Homo sapiens.
XX Key Location/Qualifiers
FH

FT CDS 871..1158
 FT /tag= a
 FT /product= Lysine-rich statherin
 PN WO200024779-A1.
 XX 04-MAY-2000.
 PD
 XX 22-OCT-1999; 99WO-US24046.
 PF
 XX 23-OCT-1998; 98US-0155209.
 PR
 XX (INCY-) INCYTE PHARM INC.
 PA
 PI Tang YT, Corley NC, Guejler KU, Patterson C;
 XX WPI: 2000-350699/30.
 DR P-PSDB; AAY94526.
 DR
 XX
 PT Purified polypeptide used for treating or preventing a disorder
 PT characterized by expression or activity of lysine-rich statherin
 PT proteins -
 XX
 XX Claim 3; Page 70; 75pp; English.
 PS
 XX The present sequence is human lysine-rich statherin protein (LRSP)
 CC cDNA from Incyte clone 2620214. This sequence was identified
 CC through analysis of a cDNA library of breast tumor tissue
 CC (BRSTW014). The LRSP sequence was found to have homology with
 CC human statherin (AAY94527) and human basic histidine-rich protein
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an
 CC inhibitor of precipitation of calcium phosphate salts in the oral
 CC cavity. The LRSP polypeptide and its antagonists may be useful for
 CC treating or preventing disorders associated with the activity of
 CC LRSP. Such disorders include autoimmune/inflammatory disorders (for
 CC example AIDS, allergies, asthma, diabetes mellitus), bacterial and
 CC fungal infection and cancers (such as leukemia, adenocarcinoma,
 CC melanoma). Antibodies to LRSP may be useful for diagnosis of
 CC the above disorders.
 CC
 XX SO Sequence 1331 BP; 397 A; 258 C; 228 G; 447 T; 1 other;
 Alignment Scores:
 Score: 1.55e-53 Length: 1331
 522.00 Matches: 95
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0
 US-09-830-244B-1 (1-95) x AAA48963 (1-1331)
 QY 1 MetTTPhehHstLysValLgYAgLysGlnHhPhelLysValThrPhetrgLurHrAsp 20
 DB 871 ATGTGGTTTCATAAAGTGGGAGACAAAACACATTTAAAGTAAGTAACTTTGGAGAGCTGAT 930
 QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysProPhenHstLysuYrG 40
 DB 931 TTGAGTAATATATAAACTCGCTCGCTCCCTTAAGAAAAAACCCTTCCACCTTACTGT 990
 QY 41 ValLleTyTLeProLeuValProLysLeuLleLleLeuPhelLeuAspLleAaPhelle 60
 DB 991 GTCATTAAATCCCTTACCTTCCAAAGTAAATTAATCTTAATTCCTGCAATTCGCTTTA 1050
 QY 61 ProLysSerLeuLleSerGlnPhenGlnAsnAsnHstTyTThHsAsnHstHrAsnH 80
 DB 1051 CCAAGAGCCTTATCCAGGCAAGTCCAGAACCAACCACTATACGACAAACCATCAACCA 1110
 QY 81 AsnThrAsnAsnLleLgrPhenAsnLleLleSerAsnCyArgThr 95
 DB 1111 AATACCAACAATATAGCTTTAATATCATCAGTAAGTCAAGAGCA 1155
 RESULT 2

AAA42355
 ID AAA42355 standard; cDNA; 221 BP.
 XX
 AC AAA42355;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:1095.
 XX
 KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cyostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; antileukemic; antiparkinsonian;
 KW anticancer; osteoprotective; antiproliferative; neurotropic; antiparasitic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune diabetes; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Huntington's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200021990-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24205.
 XX
 PR 15-OCT-1998; 98US-0104435.
 XX
 PA (GENM) GENNETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 PI WPI: 2000-317937/27.
 DR
 XX
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 PS Claim 1; Page 394; 618pp; English.
 XX
 CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiasthmatic; antileukemic; antifungal; antiviral; antiparasitic;
 CC neurotropic; antiparkinsonian; antiproliferative; neuroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention.
 XX
 SO Sequence 221 BP; 72 A; 41 C; 44 G; 64 T; 0 other;

Alignment Scores:

Pred. No.: 1 46e-15
 Score: 200.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 38.31%
 DB: 21
 Gaps: 0

US-09-830-244B-1 (1-95) x AAA42355 (1-221)

QY 1 MetTrpPheHisLysValGlyArgGlnHisPheLysValThrPheTrpGlnThrAsp 20
 DB 106 ATGTGGTTTCATAAAGTGGGAGAAACGATTTTAAGTACTTTTGGGAGACTGAT 165

QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysProPhe 36
 DB 166 TTGAGTAAATATATATAACCTGTGCTCCCTTAAGAAAAAACCCCTTC 213

RESULT 3

AAA48964
 ID AAA48964 standard; DNA: 552 BP.

AC AAA48964;

XX 06-OCT-2000 (first entry)

DE Human statherin DNA.

KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.

OS Homo sapiens.

XX Key Location/Qualifiers
 FT 73..261
 FT CDS /*tag- a

FT /product= Statherin

XX MO200024779-A1.

XX 04-MAY-2000.

XX 22-OCT-1999; 99MO-US24046.

XX 23-OCT-1998; 98US-0155209.

XX (INCY-) INCYTE PHARM INC.

PI Tang YT, Corley NC, Guegler KJ, Patterson C;

DR WPI: 2000-350699/30.

DR P-PSDB; AAY94527.

PT Purified polypeptide used for treating or preventing a disorder
 PT characterized by expression or activity of lysine-rich statherin
 PT proteins

XX Disclosure; Page 71; 75pp; English.

XX The present invention relates to human lysine-rich statherin protein
 CC (LRSP)(AAY94526). The cDNA sequence encoding this protein was identified
 CC through analysis of a cDNA library of breast tumour tissue (BRSTN014).
 CC The LRSP sequence was found to have homology with the DNA of human
 CC statherin (the present sequence) and human basic histidine-rich protein
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an
 CC inhibitor of precipitation of calcium phosphate salts in the oral cavity.
 CC The LRSP polypeptide and its antagonists may be useful for treating or
 CC preventing disorders associated with the activity of LRSP. Such
 CC disorders include autoimmune/inflammatory disorders (for example AIDS,
 CC allergies, asthma, diabetes mellitus), bacterial and fungal infection
 CC and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to

CC LRSP may be useful for diagnosis of the above disorders.

XX Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;

Alignment Scores:

Pred. No.: 1 84e-12
 Score: 178.50
 Percent Similarity: 63.77%
 Best Local Similarity: 53.62%
 Query Match: 34.20%
 DB: 21
 Gaps: 2

US-09-830-244B-1 (1-95) x AAA48964 (1-552)

QY 27 LeuValSerLeuLysLysLysLysProPheHisLysLeuTyrCysValIleTyrIleProLeu 46
 DB 128 CTGATTCATCTGAGAGAAAT-----TTTGGCTA----- 157

QY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
 DB 158 -----GATTGGAGACTGCTGTTATGGTATGCGCTTATCAGC 196

QY 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
 DB 197 CAGTTCAGAACACACACATACCCACACACATACCAACCAATACCAATATACCT 256

QY 87 PheAsnIleIleSerAsnGlyArgThr 95
 DB 257 TTTAATATCATCATGATACCTGACGAGCA 283

RESULT 4

AACT6665
 ID AACT6665 standard; cDNA: 2121 BP.

AC AACT6665;

XX 08-FEB-2001 (first entry)

DE Human ORF2220 polynucleotide sequence SEQ ID NO:4439.

XX Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
 KW vulnery; antipapillary; antipapillary; antipapillary; antipapillary;
 KW anticonvulsant; osteopathic; antipapillary; antipapillary; antipapillary;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antidiabetic;
 KW antiviral; antibacterial; antifungal; antineumatic; antidiabetic;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antidiabetic disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.

XX MO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CUBAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI: 2000-602362/57.

DR P-PSDB: AAB42456.

PT Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -

PS Claim 5: Page 3631-3632; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasorelaxant;
 CC antidiabetic; hypotensive; dermatologic; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SO Sequence 2121 BP; 617 A; 430 C; 341 G; 732 T; 1 other;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1e-11	178.50	2121	37	7	8	17	2
Percent Similarity:	63.77%						
Best Local Similarity:	53.62%						
Query Match:	34.20%						

US-09-830-244B-1 (1-95) x AAC76665 (1-2121)

OY 27 LeuValSerLeuLysLysLysPProPheHisLeuTyrCysValIleTyrIleProLeu 46
 DB 1556 CTGATTCATCTGAAGACAAAT-----TTTGCGTA----- 1585

OY 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
 DB 1586 -----GAAITGGAAAGATTGCGTATGCGTATGCGCTTATACAC 1624

OY 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
 DB 1625 CAGTTCAGACACACACACATACCCACCAACCAATACCAACCAATACCAATATACCT 1684

OY 87 PheAsnIleIleSerAsnCysArgThr 95
 DB 1685 TTTAATATCATCATCACTACGACGACA 1711

RESULT 5

AA191635

ID AA191635 standard; cDNA: 390 BP.

XX AA191635;

XX 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 11695.

KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Dymnac RT.

DR WPI; 2001-514838/56.

DR P-PSDB: AAO11704.

XX Isolated nucleic acids and polypeptides, useful for preventing

XX diagnosing and treating e.g. leukaemia, inflammation and immune

XX disorders -

XX Claim 1: SEQ ID NO 11695; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 390 BP; 127 A; 77 C; 79 G; 106 T; 1 other;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1.65e-07	136.00	390	23	3	2	0	0
Percent Similarity:	92.86%						
Best Local Similarity:	82.14%						
Query Match:	26.05%						

US-09-830-244B-1 (1-95) x AA191635 (1-390)

OY 1 MetTyrPheHisLysValGlyArgLysGlnHisPheLysValThrPheTyrPglutThrAsp 20
 DB 206 ATGTGTTTCATCAAGTCGGAGAAAGCAACATTTAAATACCTTTGGAGACTGAA 265

OY 21 LeuSerAsnAsnLysThrLeuVal 28
 DB 266 TTGAGTAATATAATAAACAATCACTC 289

RESULT 6

AAV89525

ID AAV89525 standard; cDNA: 516 BP.

XX AAV89525;

XX 15-FEB-1999 (first entry)

DE EST clone CP289.

KW Human: secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemokinesis; chemotaxis; haemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.

XX OS Homo sapiens.
 XX PN W09845436-A2.
 XX PD 15-OCT-1998.
 XX PF 10-APR-1998; 98WO-US06955.
 XX PR 10-APR-1997; 97US-0838821.
 XX PA (GENY) GENETICS INST INC.
 XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 XX PI Racie LA, Spaulding V, Treacy M;
 XX DR WPI; 1999-070077/06.
 XX
 XX New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.
 XX
 XX Claim 1; Page 245; 618pp; English.
 XX
 XX The present sequence represents a human expressed sequence tag (EST).
 CC The polynucleotide, which is a secreted EST, and the encoded protein
 CC are predicted to have useful biological activities which would make
 CC them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data is
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.
 XX
 XX Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1.33e-05 Length: 516
 Score: 121.50 Matches: 28
 Percent Similarity: 61.02% Conservative: 8
 Best Local Similarity: 47.46% Mismatches: 16
 Query Match: 23.28% Indels: 7
 DB: 20 Gaps: 2
 US-09-830-244B-1 (1-95) x AAV89525 (1-516)
 QY 30 LeuLysLysLysLysProPheHisLeuTyxCysValIleTyrllePro----- 45
 Db 41 TTGAGTAAAGAGAACCCAGCCAACTATGAGTTCCCTGTGCTTTCCTCTTGCGCTC 100
 QY 46 ---LeuValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeu 64
 Db 101 TCATGGTTTCATGATGGAGCTGATCATCTG-----AAGACTATGGTATGGCCCTT 154
 QY 65 IleserGlnPheGlnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsn 83
 Db 155 ATCAGCCAGTTCGACACACCACTATATCCACACCAACCATACCAACCATACCAAC 211
 RESULT 7
 AAT24320
 ID AAT24320 standard; cDNA to mRNA; 126 BP.
 XX AC AAT24320;
 XX DT 22-SEP-1996 (first entry)
 XX DE Human gene signature HUMGS06347.
 XX
 XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW

cell typing; abnormal cell function; ss.
 XX Homo sapiens.
 XX PN W09514772-A1.
 XX PD 01-JUN-1995.
 XX PF 11-NOV-1994; 94WO-JP01916.
 XX PR 12-NOV-1993; 93JP-0355504.
 XX PA (MATS/) MATSUBARA K.
 XX PA (OKUB/) OKUBO K.
 XX PI Matsubara K, Okubo K;
 XX DR WPI; 1995-206931/27.
 XX
 XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 XX Claim 1; Page 1582; 2245pp; Japanese.
 XX
 XX A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX
 XX Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;
 SQ
 Alignment Scores:
 Pred. No.: 7.9e-06 Length: 126
 Score: 117.00 Matches: 20
 Percent Similarity: 78.57% Conservative: 2
 Best Local Similarity: 71.43% Mismatches: 6
 Query Match: 22.41% Indels: 0
 DB: 16 Gaps: 0
 US-09-830-244B-1 (1-95) x AAT24320 (1-126)
 QY 1 MetTrpPheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
 Db 26 ATGTGTTTCATCAAGTGGGAGAGAAAGACCAATTTAAAAATANCTTTTGGGNGACTGAN 85
 QY 21 LeuSerAsnAsnLysThrLeuVal 28
 Db 86 TTGAGTAATAATAAANCTTCAGTC 109
 RESULT 8
 ABL54353
 ID ABL54353 standard; DNA; 9881 BP.
 XX AC ABL54353;
 XX DT 29-JUL-2002 (first entry)
 XX DE Chemically treated apoptosis gene #27.
 XX KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
 KW

```

XX OS Homo sapiens.
XX PN WO9845436-A2.
XX PD 15-OCT-1998.
XX PF 10-APR-1998; 98WO-US06955.
XX PR 10-APR-1997; 97US-0838821.
XX PA (GEM) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX DR WPI; 1999-070077/06.
XX PT New polynucleotides encoding human secreted proteins - derived from
XX PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX PT ovary, pituitary, retina and colon cDNA libraries.
XX PS Claim 1; Page 245; 618pp; English.
XX CC The present sequence represents a human expressed sequence tag (EST).
XX CC The polynucleotide, which is a secreted EST, and the encoded protein
XX CC are predicted to have useful biological activities which would make
XX CC them suitable for treating, preventing or ameliorating medical
XX CC conditions in humans and animals, although no supporting data is
XX CC given. Suggested activities include nutritional activity, immune
XX CC stimulating or suppressing activity, haematopoiesis regulating
XX CC activity, tissue growth activity, activin/inhibin activity,
XX CC chemotactic/chemokinetic activity, activin/inhibin activity,
XX CC activity, receptor/ligand activity, haemostatic and thrombolytic
XX CC adherin/tumour invasion suppressor activity, anti-inflammatory activity,
XX CC activity. The polynucleotide may also be useful for gene therapy.
XX SO Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other;

Alignment Scores:
Pred. No.: 1-33e-05 Length: 516
Score: 121.50 Matches: 28
Percent Similarity: 61.02% Conservative: 8
Best Local Similarity: 47.46% Mismatches: 16
Query Match: 23.28% Indels: 7
DB: 2 Gaps: 2

US-09-830-244B-1 (1-95) x AAV89525 (1-516)
QY 30 LeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIlePro----- 45
Db 41 TTGAGTAAAGAGAACCCAGCCCAATGATGAAGTTCCTGCTTGGCTTCATCTGGGCTC 100
QY 46 --LeuValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeu 64
Db 101 TCATGGTTTCATGATGGAGCTGATTCATCTG-----AAGAGTATGGGTATGGCCTT 154
QY 65 ILeSerGlnPheGlnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsn 83
Db 155 ATCAGCAGTTCAGACACACCACTATACCCACACCACTACCAACCAACCAATACCAAC 211

RESULT 7
AAT24320
ID AAT24320 standard; cDNA to mRNA; 126 BP.
XX AC AAT24320;
XX AC AAT24320;
XX DT 22-SEP-1996 (first entry)
XX DE Human gene signature HUMGS06347.
XX DE
XX DE Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX DE human; cloning; mapping; non-biased library; diagnosis; detection;
XX KW

cell typing; abnormal cell function; ss.
Homo sapiens.
WO9514772-A1.
01-JUN-1995.
11-NOV-1994; 94WO-JP01916.
12-NOV-1993; 93JP-0355504.
(MATS/) MATSUBARA K.
(OKUB/) OKUBO K.
Matsubara K, Okubo K;
WPI; 1995-206931/27.
Identifying gene signatures in 3'-directed human cDNA library - e.g.
for diagnosis of abnormal cell function, by preparing cDNA that
reflects relative abundance of corresp. mRNA in specific human
tissues
Claim 1; Page 1582; 2245pp; Japanese.
A single-stranded DNA (or its complementary strand or the corresp.
double-stranded DNA) which comprises one of the 7837 "GS" sequences
given in AAT19001-726837 and which is able to hybridise to part of
human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
sequences were obtained from 3'-directed cDNA libraries prepared
from various human tissues; synthesis of cDNA was initiated from the
3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
untranslated sequence is unique to a particular mRNA species, almost
all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
is constructed so as to reflect accurately the relative abundance of
different mRNAs in the particular tissue from which it was derived.
The appearance frequency of a given GS in a cDNA library can be
determined (esp. using primers and probes derived from the GS
sequences) as a means of diagnosing abnormal cell function or for
recognising different cell types.
SQ Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;

Alignment Scores:
Pred. No.: 7.9e-06 Length: 126
Score: 117.00 Matches: 20
Percent Similarity: 78.57% Conservative: 2
Best Local Similarity: 71.43% Mismatches: 6
Query Match: 22.41% Indels: 0
DB: 16 Gaps: 0

US-09-830-244B-1 (1-95) x AAT24320 (1-126)
QY 1 MetTrpPheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
Db 26 ATGTGGTTTCATCAAGTGGGAAGAAAGACGAATTTAAAAATANCTTTTGGGNGACTGAN 85
QY 21 LeuSerAsnAsnLysThrLeuVal 28
Db 86 TTGAGTAATAATAAANCTTCAGTC 109

RESULT 8
ABL54353
ID ABL54353 standard; DNA; 9881 BP.
XX AC ABL54353;
XX AC ABL54353;
XX DT 29-JUL-2002 (first entry)
XX DE Chemically treated apoptosis gene #27.
XX DE Apoptosis; HIV; Bloom syndrome; cardiopathy;
XX KW

```

KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
KW amyotrophic lateral sclerosis; cancer; ds.
OS Unidentified.
XX WO200177164-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-EP03969.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017444/02.
XX Chemically modified sequences of genes associated with apoptosis are
XX useful to determine methylation patterns of genomic DNA samples for
XX diagnosis of associated diseases such as cancer.
XX Claim 1; Seq ID #53; 24pp; English.
XX This invention relates to chemically pre-treated DNA of genes
XX associated with apoptosis. The nucleic acids are used to allocate
XX patients for specific therapy for HIV infection, Bloom syndrome,
XX cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
XX infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
XX and cancer. This nucleotide sequence represents a chemically
XX treated apoptosis gene. Even SEQ ID numbers are the complementary
XX DNA strands to the odd SEQ ID numbers. The sequence data for this
XX patent is not represented in the printed specification but is based on
XX information supplied by the European patent office.
XX Sequence 9881 BP; 2606 A; 196 C; 2308 G; 4771 T; 0 other;

Alignment Scores:
Pred. No.: 0.00221 Length: 9881
Score: 116.50 Matches: 28
Percent Similarity: 51.32% Conservative: 11
Best Local Similarity: 36.84% Mismatches: 15
Query Match: 22.32% Indels: 22
DB: 24 Gaps: 1
US-09-830-244B-1 (1-95) x ABL54353 (1-9881)
QY 1 MetTppPheHisLysValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAsp 20
Db 5782 ATGTGTTTATTAAGTCGGGAAGAAGTAGTAATTAATAAATAATTTTGGGAGATGAA 5841
QY 21 LeuSerAsnAsnLysThrLeuValSerLeu----- 30
Db 5842 TTGAGTAATAATAAATAATTTAGTTTTCGTTAATAATAATAATAATAATAATAA 5901
QY 31 -----LysLysLysLysPropheHisLeu-Ty 39
Db 5902 TAATTTATGTAATGTTAGTTTATGATGAGGAAAGGAGGATTTTGTATTT 5961
QY 39 rCysValIleThrProLeuValProLysLeuIleLeuPhe 54
Db 5962 TGTGTTAGTTTATTAAGTATTAATAAGTTAATAGTTTGTGTT 6007
RESULT 9
ABN43273
ID ABN43273 standard; DNA; 60 BP.
XX
XX
XX AC ABN43273;
XX

DT 15-JUL-2002 (first entry)
XX Human spliced transcript detection oligonucleotide SEQ ID NO:16021.
DE Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
KW Homo sapiens.
XX WO200210449-A2.
XX 07-FEB-2002.
XX 20-JUL-2001; 2001WO-IB01903.
XX 28-JUL-2000; 2000US-221607P.
XX 02-MAY-2001; 2001US-287724P.
XX (COMP-) COMPUGEN INC.
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX WPI; 2002-257383/30.
XX New oligonucleotide libraries comprising oligonucleotides which
XX selectively hybridize to mRNAs transcribed from a transcription unit of
XX a genome, useful for detecting tissue-, pathology-, and
XX developmental-specific genes
XX Example 1; SEQ ID 16021; 47pp; English.
XX The present invention describes oligonucleotide libraries for detecting
XX messenger RNAs that populate a (sub-)transcriptome, where the
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX transcription units that populate a genome. The library comprises
XX several oligonucleotides, each capable of hybridizing selectively to a
XX set of messenger RNAs transcribed from a given transcription unit of
XX the genome, which encodes one or more messenger RNA splice variants.
XX The oligonucleotide libraries are useful for detecting mRNAs from a
XX biological sample, in expression profiling studies, in qualitatively or
XX quantitatively characterizing the corresponding transcriptome, and in
XX detecting RNA transcripts and splice variants of human or animal
XX transcriptomes. The libraries may also be used as specialised mini
XX libraries to detect transcripts of a sub-transcriptome under a
XX particular biological or pathological state, and so allowing the
XX detection of tissue- and pathology-specific genes such as those genes
XX only expressed in specific tissue under a specific pathological
XX condition; to detect developmental specific genes; and to detect RNA
XX transcripts and splice variants of a transcriptome of a patient suffering
XX from a particular disorder. ABN27253 to ABN59389 represent
XX oligonucleotide sequences from rats, humans and mice, which are used in
XX the exemplification of the present invention.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 60 BP; 22 A; 25 C; 4 G; 9 T; 0 other;

Alignment Scores:
Pred. No.: 9.45e-06 Length: 60
Score: 113.00 Matches: 19
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 0
Query Match: 21.65% Indels: 0
DB: 24 Gaps: 0
US-09-830-244B-1 (1-95) x ABN43273 (1-60)
QY 63 SerLeuIleSerGlnPheGlnAsnAsnHisTyThrHisAsnHisThrAsnHisAsnThr 82
Db 1 GCCCTTATCAGCCAGTTCACAGAACACCACTATACCAACCAACCAATACC 60
RESULT 10

ABL92199/c
 ID ABL92199 standard; DNA; 6070 BP.
 AC ABL92199;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated DNA repair gene fragment complementary to#4.
 XX
 KW DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3;
 KW PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;
 KW DITL1; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;
 KW Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;
 KW immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;
 KW cancer; ds.
 XX
 OS Unidentified.
 XX
 PN WO200181622-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03972.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-034446/04.
 XX
 PT New nucleic acid derived from genes associated with DNA repair, useful
 PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
 PT cytosine methylation.
 XX
 PS Claim 1; SEQ ID NO 8; 25pp + sequence listing; English.
 XX
 CC The invention relates to nucleic acids containing a sequence of at least
 CC 18 nucleotides of chemically treated DNA of genes associated with DNA
 CC repair, and their complements. The invention also relates to nucleic
 CC acids comprising at least 18 base pairs of the chemically pretreated DNA
 CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,
 CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDT1, TDG, INPPL1,
 CC RFC4, DITL1, FANCB, or XRCC8. Nucleic acids of the invention and related
 CC oligomers, are useful for diagnosis of diseases associated with gene
 CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
 CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
 CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours
 CC and cancer, particularly by determining status of cytosine methylation
 CC and/or by detecting single-nucleotide polymorphisms. Determination of
 CC individual methylation patterns may allow development of individualised
 CC therapies. The sequences given in records ABL92192-ABL92335 represent
 CC chemically pre-treated DNA fragments from genes associated with DNA
 CC repair, and their complements.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 6070 BP; 1649 A; 144 C; 1328 G; 2949 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.0897 Length: 6070
 Score: 101.00 Matches: 33
 Percent Similarity: 47.19% Conservative: 9
 Best Local Similarity: 37.08% Mismatches: 33
 Query Match: 19.35% Indels: 14
 DB: 24 Gaps: 3

QY 19 ThrAspLeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysProPheHisLeu 38
 III III III III III III III III III III III III III III III
 Db 675 ACCCAACCTCTAACACTACTACTCGTATCCCTCTATAAACAACGAAACACTTTA 616
 QY 39 TyrCysValIleTyrIleProLeuValProLysLeuIleLeuLeuPheLeuAspIleAla 58
 III III III III III III III III III III III III III III III
 Db 615 TACTTTTCTTATACCTTCCATAATACGAACTAACTACTCTTTATTTTAAATATCTAC 556
 QY 59 PheIleProLysSer-----LeuIleSerGlnPhe 68
 ::::: III III III III III III III III III III III III III III
 Db 555 TATATACCATATTTCTTAACACCCCTTAATCTTACAACTATACTTACGATATCCCAA--- 499
 QY 69 GlnAsn-----AsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIle 85
 III III III III III III III III III III III III III III III
 Db 498 CAAACGAAATACTTAAATAATCTACTACATCAATTAATTAATTCGAATTCCTACTCTTC 439
 QY 86 ArgPheAsnIleLeuSerAsnCysArg 94
 III III III III III III III III III III III III III III III
 Db 438 ACATTTAATCTATATCTACTACTACAGT 412
 RESULT 11
 ABL49310/c
 ID ABL49310 standard; DNA; 6070 BP.
 XX
 AC ABL49310;
 XX
 DT 01-MAY-2002 (first entry)
 XX
 XX Human polynucleotide associated with DNA replication SEQ ID NO 10.
 DE
 DE Human; cytostatic; neuroprotective; nootropic; immunostimulant;
 KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
 KW ERCC3; SNRPA1; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;
 KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200177377-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03971.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-017471/02.
 XX
 PT New nucleic acid sequences from chemically modified genes associated
 PT with DNA replication, useful for analysing cytosine methylations for
 PT diagnosis and therapy of diseases e.g. Ataxia telangiectasia
 XX
 PS Claim 1; SEQ ID NO 10; 23pp + Sequence Listing; English.
 XX
 CC The invention relates to nucleic acid sequences comprising at least 18
 CC bases of a chemically pretreated gene associated with gene regulation,
 CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
 CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPA1,
 CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases
 CC unethylated at the 5-position to uracil or another base with
 CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
 CC cytosine methylations. The DNA sequences and method are useful in the
 CC diagnosis of diseases (or predisposition to diseases) associated with the
 CC replication and in therapy of such diseases, by enabling analysis of the
 CC cytosine methylation patterns of such genes. They are especially useful
 CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's

QY 3 PheHisLysValGlyArgLysGlnHisPheLys-ValThrPheTrpGluThrAspLeuSe 22
 Db 4931 TTTAAAACTAAACAAAAAACCACTTAAACCAACGTC---GAAAAAACCTTAA 4875
 QY 22 rAsnAsn-----LysThrLeuValSerLeuLysLysLysPr 35
 Db 4874 CAACAATAAACAATATCCATCTCTACAAACATTTAAATAATATAAAAAA 4815
 QY 35 o-----PheHisLeuTyrCy 40
 Db 4814 AAAAAACCTACCTCTACAAATATCAATAAATAAATAAATAATATACATACAAAAA 4755
 QY 40 sValileTyr-----IleProLeuValProLysLeuLeileLeuPheLeuAs 56
 Db 4754 CTTAAATATATTTCAAAAAAATAAATAATTTAACTTAAACCAACACACAAACCTCTCCA 4695
 QY 56 pIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsn----- 70
 Db 4694 ATTAACCTCAATTCCTCCCATATCTCTACATCTCTTCAAAATCCAAAAAACGTTAA 4635
 QY 71 -----AsnHisTyrThrHisAsnHisThrAsnHisAsnth 82
 Db 4634 ACGAATCTTTAATTTCAAAAAAATAAATAATTTTAACTTAAACCAACCACTTAAAC 4875
 QY 82 rAsnAsnIleArgPhe 87
 Db 4574 GAAAAACCTCTCTCTTC 4559

RESULT 15

ABL34097/c

ID ABL34097 standard; DNA; 7038 BP.

XX AC ABL34097;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 2070.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anemic; cytosine methylation; antiasthmatic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 XX gene; ds.

XX OS Homo sapiens.

XX PN W0200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07537.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043626.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130909/17.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX

SQ Sequence 7038 BP; 1424 A; 375 C; 2062 G; 3177 T; 0 other;

Alignment Scores:

Pred. No.: 1.52 Length: 7038
 Score: 91.50 Matches: 37
 Percent Similarity: 38.89% Conservative: 12
 Best Local Similarity: 29.37% Mismatches: 35
 Query Match: 17.53% Indels: 42
 DB: 24 Gaps: 5

US-09-830-244B-1 (1-95) x ABL34097 (1-7038)

QY 3 PheHisLysValGlyArgLysGlnHisPheLys-ValThrPheTrpGluThrAspLeuSe 22
 Db 4931 TTTAAAACTAAACAAAAAACCACTTAAACCAACGTC---GAAAAAACCTTAA 4875
 QY 22 rAsnAsn-----LysThrLeuValSerLeuLysLysLysPr 35
 Db 4874 CAACAATAAACAATATCCATCTCTACAAACATTTAAATAATATAAAAAA 4815
 QY 35 o-----PheHisLeuTyrCy 40
 Db 4814 AAAAAACCTACCTCTACAAATATCAATAAATAAATAAATAATATACATACAAAAA 4755
 QY 40 sValileTyr-----IleProLeuValProLysLeuLeileLeuPheLeuAs 56
 Db 4754 CTTAAATATATTTCAAAAAAATAAATAATTTAACTTAAACCAACACACAAACCTCTCCA 4695
 QY 56 pIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsn----- 70
 Db 4694 ATTAACCTCAATTCCTCCCATATCTCTACATCTCTTCAAAATCCAAAAAACGTTAA 4635
 QY 71 -----AsnHisTyrThrHisAsnHisThrAsnHisAsnth 82
 Db 4634 ACGAATCTTTAATTTCAAAAAAATAAATAATTTTAACTTAAACCAACCACTTAAAC 4875
 QY 82 rAsnAsnIleArgPhe 87
 Db 4574 GAAAAACCTCTCTCTTC 4559

Search completed: April 14, 2003, 23:13:50
 Job time : 232 secs

Nucleic acid comprising fragment of chemically modified gene, useful
 for diagnosis and treatment of diseases associated with abnormal
 cytosine methylation

Claim 1; SEQ ID NO 2070; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated
 genes which are modified by the methylation of cytosines. The sequences

FILING DATE: 22-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: CARTY, CHRISTINE E.
 REGISTRATION NUMBER: 36,099
 REFERENCE/DOCKET NUMBER: 19307 PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-6734
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8010 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 PCT-US95-11859-2

Alignment Scores:
 Pred. No.: 88.1 Length: 8010
 Score: 70.50 Matches: 25
 Percent Similarity: 43.88% Conservative: 18
 Best Local Similarity: 25.51% Mismatches: 31
 Query Match: 13.51% Indels: 24
 DB: 5 Gaps: 5

US-09-830-244B-1 (1-95) x PCT-US95-11859-2 (1-8010)

QY 3 PheHisLysValGlyArg-----LysGlnHisPheLysValThrPheTrpGluThrAsp 20
 DB 7717 TGGCACAATAATATAGATTAAACAGGCAACACGTTGTGAACAAATGGTGAAGTGTAT 7658
 QY 21 LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysLysProPheHisLeuTyrCys 40
 DB 7657 GCCAAGGCCACCGGAACGGTGGCAATATAAATA-----7622
 QY 41 ValIleTyrIleProLeuValProLysLeuLeuLeuPheLeuPheLeuAspIleAlaPheIle 60
 DB 7621 -----TATATT---ATTGTAACAAATAATATATATTGTGCATGTATATTACTTATGTG 7571
 QY 61 ProLysSerLeuIleSerGlnPheGln---AsnAsnHis-----72
 DB 7570 TGTAGGCGCGTGCACCAACAAACGCGTGCACACAGCCACTGAGTCACAGGCGTAGTTAA 7511
 QY 73 -----TyrThrHisAsnHisThr-AsnHisAsnThrAsnAsnIleArgPhe 87
 DB 7510 TTGTTATTGCACATAAACAACACATACATACATACAAATAAATTACATTAC 7459

RESULT 10
 US-08-998-416-915
 : Sequence 915, Application US/08998416
 : Patent No. 6239264
 : GENERAL INFORMATION:
 : APPLICANT: Philippsen, Peter
 : APPLICANT: Pohlmann, Rainer
 : APPLICANT: Steiner, Sabine
 : APPLICANT: Mohr, Christine
 : APPLICANT: Wendland, Jurgen
 : APPLICANT: Knechtle, Philipp
 : APPLICANT: Reibschung, Corinne
 : TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
 : NUMBER OF INVENTION: AND USES THEREOF
 : NUMBER OF SEQUENCES: 1152
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 6239264artis Corporation
 : STREET: 3054 Cornwallis Road
 : CITY: Research Triangle Park
 : STATE: No. 6239264th Carolina
 : COUNTRY: USA
 : ZIP: 27709
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 915:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 688 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1570RP

US-08-998-416-915
 Alignment Scores:
 Pred. No.: 5.33 Length: 688
 Score: 69.00 Matches: 24
 Percent Similarity: 42.68% Conservative: 11
 Best Local Similarity: 29.27% Mismatches: 41
 Query Match: 13.22% Indels: 6
 DB: 4 Gaps: 2

US-09-830-244B-1 (1-95) x US-08-998-416-915 (1-688)
 QY 9 LysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuVal 28
 DB 38 AAAGACAGATTCCACAGTCCACACACTTGCACAAACCATCTGTAAGCTTTCACAGATTGCGC 97
 QY 29 SerLeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeuValPro 48
 DB 98 AGCAGCAAAACCCAGCCACCGCTCTCTCAATCTCATCTCCTCAGCAACAACAGGTTCCAC 157
 QY 49 LysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSerGlnPhe 68
 DB 158 AAGCTCAGCAGCAGCTACTTCAACAACAAGGTTCCACAGCTCAGCAGCAGTACTTTC 217
 QY 69 GlnAsnAsn-----HisTyrThrHisAsnHisThrAsn---HisAsnThr 82
 DB 218 AACAAACAACAACAACAGCAACACCACCACCACCACCACCACCACCACCACCACCAACAACAACAAC 277
 QY 83 AsnAsn 84
 DB 278 CACAAC 283

RESULT 11
 US-09-272-114-2/c
 : Sequence 2, Application US/09272114
 : Patent No. 6133013
 : GENERAL INFORMATION:
 : APPLICANT: Viola, Ronald E.
 : APPLICANT: Jayasekera, Maithri M. K.
 : APPLICANT: Saribas, Abdullah S.
 : TITLE OF INVENTION: Truncated Aspartase Enzyme Derivatives
 : TITLE OF INVENTION: and Uses Thereof
 : NUMBER OF SEQUENCES: 2
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Renner, Kenner, Greive, Bobak,
 : ADDRESSEE: Taylor & Weber
 : STREET: 1610 First National Tower
 : CITY: Akron
 : STATE: Ohio


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/869,933
; FILING DATE: 19920416
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/154 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11298 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; STRAIN: FCRI beta
; US-07-869-933-31

```

```

Alignment Scores:
Pred. No.: 215          Length: 11298
Score: 69.00           Matches: 27
Percent Similarity: 40.00% Conservative: 7
Best Local Similarity: 31.76% Mismatches: 23
Query Match: 13.22% Indels: 28
DB: 1 Gaps: 3

```

```

US-09-830-244B-1 (1-95) x US-07-869-933-31 (1-11298)

```

```

QY 7 GlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnLys--- 25
Db 5669 GGTGAGAAACAGCATCATCTACTACATTTCTTA-----TTGACACGACACAAATGA 5619
QY 26 -----ThrLeuValSerLeuLysLysLysLysPropheHisLeuTyrCys 40
Db 5618 CAACAAAAAGATGCATTGGTA----- 5595
QY 41 ValIleTyrIleProLeuValProLysLeuLeuLeuPheLeuAspIleAlaPheIle 60
Db 5594 -----ITCGCCCCAAAGATTATTTCTTATAGCATCTCCTCTTTTCATC 5550
QY 61 -ProLysSerLeuIleSerGlnPheGlnAsnHis-TyrThrHisAsnHisThrAsnH 80
Db 5549 TCCTTACCCTGACGACGCTTTTAAAGTGACACATACACACACACACACACACA 5490
QY 80 isAsnThrAsn 83
Db 5489 CACACACAAAC 5479

```

```

RESULT 14

```

```

US-08-201-879A-2/c
; Sequence 2, Application US/08201879A
; Patent No. 5807988
; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; APPLICANT: JOUVIN, Marie-Helene
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN E
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/201,879A
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/869,933
; FILING DATE: 16-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03419
; FILING DATE: 16-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/234/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11298 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: FCRI beta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(456..511, 1381..1510, 2026..2160, 4475..4531,
; US-08-201-879A-2

```

```

Alignment Scores:
Pred. No.: 215          Length: 11298
Score: 69.00           Matches: 27
Percent Similarity: 40.00% Conservative: 7
Best Local Similarity: 31.76% Mismatches: 23
Query Match: 13.22% Indels: 28
DB: 1 Gaps: 3

```

```

US-09-830-244B-1 (1-95) x US-08-201-879A-2 (1-11298)

```

```

QY 7 GlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnLys--- 25
Db 5669 GGTGAGAAACAGCATCATCTACTACATTTCTTA-----TTGACACGACACAAATGA 5619
QY 26 -----ThrLeuValSerLeuLysLysLysLysPropheHisLeuTyrCys 40
Db 5618 CAACAAAAAGATGCATTGGTA----- 5595
QY 41 ValIleTyrIleProLeuValProLysLeuLeuLeuPheLeuAspIleAlaPheIle 60
Db 5594 -----TTCGCCCCAAAGATTATTTCTTATAGCATCTCCTCTTTTCATC 5550
QY 61 -ProLysSerLeuIleSerGlnPheGlnAsnHis-TyrThrHisAsnHisThrAsnH 80
Db 5549 TCCTTACCCTGACGACGCTTTTAAAGTGACACATACACACACACACACACA 5490
QY 80 isAsnThrAsn 83
Db 5489 CACACACAAAC 5479

```

```

RESULT 15
US-09-103-663-31/c
; Sequence 31, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:

```

APPLICANT: Kinet et al.
TITLE OF INVENTION: Isolation, characterization, and use of the human beta
TITLE OF INVENTION: subunit of the high affinity receptor for
TITLE OF INVENTION: immunoglobulin E.
FILE REFERENCE: 50490
CURRENT APPLICATION NUMBER: US/09/103,663D
CURRENT FILING DATE: 1998-06-23
EARLIER APPLICATION NUMBER: 07/869,933
EARLIER FILING DATE: 1992-04-16
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 31
LENGTH: 11298
TYPE: DNA
ORGANISM: Homo sapiens
US-09-103-663-31

Alignment Scores:
Pred. NO.: 215 Length: 11298
Score: 69.00 Matches: 27
Percent Similarity: 40.00% Conservative: 7
Best Local Similarity: 31.76% Mismatches: 23
Query Match: 13.22% Indels: 28
DB: 4 Gaps: 3

US-09-830-244B-1 (1-95) x US-09-103-663-31 (1-11298)

```
Oy 7 GlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnAsnLys--- 25
   ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5669 GGTGAGAAACAGCATCATCATCAATTCCTA-----TTGAACAGCAACAATGA 5619

Oy 26 -----ThrLeuValSerLeuLysLysLysLysLysProPheHisLeuTyrCys 40
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5618 CAAACAAAAGAGTCACATTGTA----- 5595

Oy 41 ValIleTyrIleProLeuValProLysLeuIleIleLeuPheLeuAspIleAlaPheIle 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5594 -----TTGCGCCCAAAAGTTATTCTTATAGCATCTCCTGTTTCATC 5550

Oy 61 -ProLysSerLeuIleSerGlnPheGlnAsnAsnHis-TyrThrHisAsnHisThrAsnH 80
   ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5549 TCCCTACCATCTGACCAAGTCCTTTTAAAGTGACACATACACACACACACACACACA 5490

Oy 80 IsAsnThrAsn 83
   ::||| |||||
Db 5489 CACACACAAC 5479
```

Search completed: April 15, 2003, 00:03:56
Job time : 104 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2003, 23:09:51 ; Search time 85 Seconds
(without alignments)

Title: US-09-830-244B-1
Perfect score: 522
Sequence: 1 MWFKVGRKQHKFTWETD.....NHTNHTNNIRFNISNCR 95

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 593429 seqs, 439583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09830244/runat_10042003_090017_27901/app_query.fasta_1.263
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09830244 -CGEN_1_1.80 -runat_10042003_090017_27901
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_XMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	83.5	16.0	12571	10	US-09-965-553-20
C 2	79	15.1	2000	9	US-09-938-842A-4834
C 3	79	15.1	56737	10	US-09-782-378A-17
C 4	77.5	14.8	5109	9	US-09-422-569-11
C 5	77.5	14.8	5109	9	US-09-422-569-11

5	75	14.4	1503841	9	US-09-946-807-1	Sequence 1, Appli
6	75	14.4	1503841	10	US-09-795-668-1	Sequence 1, Appli
7	75	14.4	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 8	74.5	14.3	32038	9	US-10-079-854-292	Sequence 292, App
C 9	74.5	14.3	32038	10	US-09-764-878-292	Sequence 292, App
10	73	14.0	1248	10	US-09-841-132-358	Sequence 358, App
11	73	14.0	1278	10	US-09-841-132-358	Sequence 358, App
12	73	14.0	1311	10	US-09-841-132-359	Sequence 359, App
13	72	13.8	6700	9	US-09-938-842A-4705	Sequence 4705, Ap
14	72	13.8	6700	10	US-09-759-152-3	Sequence 3, Appli
15	72	13.8	8161	10	US-09-759-152-9	Sequence 9, Appli
16	72	13.8	8175	10	US-09-759-152-7	Sequence 7, Appli
17	72	13.8	8518	10	US-09-759-152-5	Sequence 5, Appli
C 18	71	13.6	2000	9	US-09-938-842A-3444	Sequence 3444, Ap
19	70.5	13.5	151	10	US-09-864-761-22646	Sequence 22646, A
20	70.5	13.5	492	10	US-09-864-761-5885	Sequence 5885, Ap
21	70.5	13.5	99014	10	US-09-880-107-3428	Sequence 3428, Ap
22	70.5	13.5	465237	10	US-09-933-267A-1	Sequence 1, Appli
C 23	70	13.4	239	10	US-09-741-669-14	Sequence 14, Appli
C 24	70	13.4	2563	9	US-10-072-349-250	Sequence 250, App
C 25	70	13.4	2563	9	US-10-072-349-251	Sequence 251, App
C 26	70	13.4	2563	10	US-09-764-855-250	Sequence 250, App
C 27	70	13.4	2563	10	US-09-764-855-251	Sequence 251, App
C 28	70	13.4	7107	10	US-09-070-927A-527	Sequence 527, App
29	69.5	13.3	943	10	US-09-974-300-2072	Sequence 2072, Ap
C 30	69.5	13.3	1200	10	US-09-887-576-722	Sequence 722, App
C 31	69.5	13.3	2000	9	US-09-938-842A-2853	Sequence 2853, Ap
32	69	13.2	2886	10	US-09-801-368-131	Sequence 131, App
33	69	13.2	12541	10	US-09-764-877-2184	Sequence 2184, Ap
34	68.5	13.1	435	10	US-09-864-761-16771	Sequence 16771, A
35	68.5	13.1	458	10	US-09-770-444-474	Sequence 474, App
C 36	68.5	13.1	1863	10	US-09-765-873A-13	Sequence 13, Appli
C 37	68.5	13.1	2307	9	US-10-278-173-160	Sequence 160, App
C 38	68.5	13.1	2658	9	US-09-858-525A-5	Sequence 5, Appli
C 39	68.5	13.1	4325	9	US-09-858-525A-6	Sequence 6, Appli
40	68.5	13.1	30310	10	US-09-800-631-96	Sequence 96, Appli
C 41	68.5	13.1	152331	9	US-10-095-407-17	Sequence 17, Appli
42	68.5	13.1	176373	9	US-10-095-407-17	Sequence 17, Appli
C 43	68.5	13.1	465237	10	US-09-933-267A-1	Sequence 1, Appli
C 44	68	13.0	721	10	US-09-917-800A-881	Sequence 881, App
C 45	67.5	12.9	599	9	US-09-796-692-8191	Sequence 8191, Ap

ALIGNMENTS

RESULT 1
US-09-965-553-20/C
; Sequence 20, Application US/09965553
; Patent No. US20020112259A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/965,553
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087125
; PRIOR FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 20
; LENGTH: 12571
; TYPE: DNA
; ORGANISM: Glycine max
US-09-965-553-20

Alignment Scores:
Pred. No.: 10.9
Score: 83.50
Percent Similarity: 44.29%
Best Local Similarity: 31.43%
Length: 12571
Matches: 22
Conservative: 9
Mismatch: 28

; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 5109
 ; TYPE: DNA
 ; ORGANISM: Caenorhabditis elegans
 US-09-422-569-11

Alignment Scores:

Pred. No.: 16.8 Length: 5109
 Score: 77.50 Matches: 29
 Percent Similarity: 40.00% Conservative: 13
 Best Local Similarity: 27.62% Mismatches: 24
 Query Match: 14.85% Indels: 39
 DB: 9 Gaps: 5

US-09-830-244B-1 (1-95) x US-09-422-569-11 (1-5109)

QY 11 HisPhelysValThrPheTrpGluThrAspLeuSerAsnAsnLys----- 25
 Db 1530 CATCAAAATTTGGTGTCTTTTGCAGAAAGATGTGGTTCAAACAAAGGTGCAATCATTTGA 1589
 QY 26 -----ThrLeuValSerLeuLys 31
 Db 1590 CTCATGTGGCGGTGTGTTCATAGCGACAGTGCATCATCCTTGGTGTGCTGAAG 1649
 QY 32 LysLysLysProPheHisLeuTyrCysValIleTyrIleProLeuValProLysLeuIle 51
 Db 1650 AAGAAACAG-----TACACATCCATTCATCATGTGTGGTGGAG----- 1688
 QY 52 IleLeuPheLeuAspIleAlaPheIleProLysSerLeu---IleSerGlnPheGlnAsn 70
 Db 1689 -----GTTGCGCGCTGTCCACCCAGAGGAGCGCCACCTGTCCAAGATGCCAGCAG 1739
 QY 71 AsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArgPheAsnIleIle 90
 Db 1740 AAGCGGTAC-----GAAATCCAACTACATAATCTTTG 1772
 QY 91 SerAsnCysArgThr 95
 Db 1773 AGCAGATGCAGAACT 1787

RESULT 5

US-09-946-807-1
 ; Sequence 1, Application US/09946807
 ; Patent No. US20020165144A1
 ; GENERAL INFORMATION:

; APPLICANT: Stefansson, Hreinn
 ; APPLICANT: Steinthorsdottir, Valgerdur
 ; APPLICANT: Gulcher, Jeffrey R.
 ; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
 ; FILE REFERENCE: 2345.2004-001
 ; CURRENT APPLICATION NUMBER: US/09/946,807
 ; CURRENT FILING DATE: 2001-09-05
 ; PRIOR APPLICATION NUMBER: US/09/795,668
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 09/515,716
 ; PRIOR FILING DATE: 2000-02-28
 ; NUMBER OF SEQ ID NOS: 1531
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1

; LENGTH: 1503841
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens

; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: r-g or a
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: y-t/u or c
 ; FEATURE:

; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: m-a or c
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: k-g or t/u
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: s-g or c
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: w-a or t/u
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: b-g or c or t/u
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: d-a or g or t/u
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: h-a or c or t/u
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: v-a or g or c
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(1531)
 ; OTHER INFORMATION: n-a or g or c or t/u
 ; OTHER INFORMATION: n-a or g or c or t/u
 US-09-946-807-1

Alignment Scores:

Pred. No.: 1.46e+05 Length: 1503841
 Score: 75.00 Matches: 21
 Percent Similarity: 51.02% Conservative: 4
 Best Local Similarity: 42.86% Mismatches: 18
 Query Match: 14.37% Indels: 6
 DB: 9 Gaps: 3

US-09-830-244B-1 (1-95) x US-09-946-807-1 (1-1503841)

QY 43 TyrIleProLeuValProLysLeuIleIleLeu-----PheLeuAspIleAlaPhe 59
 Db 239178 TACATCCCATTCACCCCTCCATTTTATATCATCTCTCTCTT-----TGTTTC 239231
 QY 60 IleProLysSerLeuIleSerGlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsn 79
 Db 239232 TTCTCCTCAAGTCTCTCTCTCTCTCTCTTACAAAACACACACACACACACACACAC 239288
 QY 80 HisAsnThrAsnAsnIleArgPheAsn 88
 Db 239289 ACACACACACACACACGACGACAAAGACAAAC 239315

RESULT 6

US-09-795-668-1
 ; Sequence 1, Application US/09795668
 ; Patent No. US20020045577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stefansson, Hreinn
 ; APPLICANT: Steinthorsdottir, Valgerdur
 ; APPLICANT: Gulcher, Jeffrey R.
 ; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
 ; FILE REFERENCE: 2345.2004-001
 ; CURRENT APPLICATION NUMBER: US/09/795,668
 ; CURRENT FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 09/515,716
 ; PRIOR FILING DATE: 2000-02-28

; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 292
; LENGTH: 32038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-292

```

Alignment Scores:

```

Pred. No.: 612          Length: 32038
Score: 74.50           Matches: 26
Percent Similarity: 32.58% Conservative: 17
Best Local Similarity: 19.70% Mismatches: 48
Query Match: 14.27% Indels: 41
DB: 9                  Gaps: 4

```

US-09-830-244B-1 (1-95) x US-10-079-854-292 (1-32038)

```

QY 3 PheHisLysValGlyArgLysGlnHis---PheLysValThrPheTrpGluThrAsp--- 20
Db 25488 TTTTCATAAACAGCCAGGAGTACTCATTTTAAATAATCCATTTAAGAGCACTCCAGAC 25429
QY 21 -----LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysPro 35
Db 25428 AGAAGATGTTAATGAATTAATCCCAAGAGCTCTATGCTCTCTCATGAAGAAT 25369
QY 36 PheHisLeuTyrCysValIleTyrIleProLeuValProLysLeuIleLeuPheLeu 55
Db 25368 TAAAGGATTATATGTTCTCTTTATGCCACCATTTCAACCTTGCTGCTACTCAT 25309
QY 56 AspIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsnAsnHisTyrThrHis 75
Db 25308 GACAGCACCTACTTCTCCGCTTTTATATATATGCTAATACTTTGGCAAGTATATGA 25249
QY 76 Asn-----
Db 25248 TCTCCCTGATGTATCTGTAGGCTCTAAACACTATGTGGAATGCTGATAATACTAA 25189
QY 77 -----HisThrAsnHisAsnThr 82
Db 25188 TAAGGATACAGTAAATGAGGCTCTAAACACTATGTGGAATGCTGATAATACTAA 25132
QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArg 94
Db 25131 AATAATTAGTACATAATGTTAATCAACTGCTCAG 25096

```

RESULT 9

```

US-09-764-878-292/c
; Sequence 292, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 292
; LENGTH: 32038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-292

```

Alignment Scores:

```

Pred. No.: 612          Length: 32038
Score: 74.50           Matches: 26
Percent Similarity: 32.58% Conservative: 17
Best Local Similarity: 19.70% Mismatches: 48
Query Match: 14.27% Indels: 41
DB: 10                 Gaps: 4

```

US-09-830-244B-1 (1-95) x US-09-764-878-292 (1-32038)

```

QY 3 PheHisLysValGlyArgLysGlnHis---PheLysValThrPheTrpGluThrAsp--- 20
Db 25488 TTTTCATAAACAGCCAGGAGTACTCATTTTAAATAATCCATTTAAGAGCACTCCAGAC 25429
QY 21 -----LeuSerAsnAsnLysThrLeuValSerLeuLysLysLysPro 35
Db 25428 AGAAGATGTTAATGAATTAATCCCAAGAGCTCTATGCTCTCTCATGAAGAAT 25369
QY 36 PheHisLeuTyrCysValIleTyrIleProLeuValProLysLeuIleLeuPheLeu 55
Db 25368 TAAAGGATTATATGTTCTCTTTATGCCACCATTTCAACCTTGCTGCTACTCAT 25309
QY 56 AspIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsnAsnHisTyrThrHis 75
Db 25308 GACAGCACCTACTTCTCCGCTTTTATATATATGCTAATACTTTGGCAAGTATATGA 25249
QY 76 Asn-----
Db 25248 TCTCCCTGATGTATCTGTAGGCTCTAAACACTATGTGGAATGCTGATAATACTAA 25189
QY 77 -----HisThrAsnHisAsnThr 82
Db 25188 TAAGGATACAGTAAATGAGGCTCTAAACACTATGTGGAATGCTGATAATACTAA 25132
QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArg 94
Db 25131 AATAATTAGTACATAATGTTAATCAACTGCTCAG 25096

```

RESULT 10

```

US-09-841-132-358
; Sequence 358, Application US/09841132
; Patent No. US20020051848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Rasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841.132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 358
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-841-132-358

```

Alignment Scores:

```

Pred. No.: 7.88          Length: 1248
Score: 73.00           Matches: 30
Percent Similarity: 43.36% Conservative: 19
Best Local Similarity: 26.55% Mismatches: 34
Query Match: 13.98% Indels: 30
DB: 10                 Gaps: 7

```

US-09-830-244B-1 (1-95) x US-09-841-132-358 (1-1248)

```

QY 2 TrpPheHis-----LysValGlyArg-----LysGlnHisPheLys 13
Db 286 TGGGAACACACTAAACAGCTGGCGATATTCCTACTTTTGGAAAGCTATCTTTTCCA 345
QY 14 ValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuValSerLeuLysLys 33

```

```

Db 346 GCCTCT-----TCTTCTCGGAATCCTTATTGAACCTCAAGAACCC 387
Qy 34 LysProPheHisLeu-----TyrCysValIleTyrIleProLeuVal 47
Db 388 GAGCCTCAACTATTGGCGATATTAGCCTCTCGGTTTTCGTGTGTCGTCAGAAAT 447
Qy 48 ProLysLeuIleLeuPheLeuAspIleAala---PheIleProLysSerLeuIleSer 66
Db 448 CCT-----TTTCTTTCTCTGGACCTTTTATGCCAAAAACCTATGTGCAA 492
Qy 67 -----GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
Db 493 GGGCAAAACGCTGCTTCTACAAAAAACCTTTATTACTATGACCATGCGCATGTGGAATTA 552
Qy 83 AsnAsnIleArgPheAsnIleLeuSerAsnCysArgThr 95
Db 553 CATTCATAGACTTTCGCATCATTCCTCCACATTTACACA 591

RESULT 11
US-09-841-132-538
; Sequence 538, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 538
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: C. Trachomatis D serovar
US-09-841-132-538

Alignment Scores:
Pred. No.: 8.17 Length: 1278
Score: 73.00 Matches: 30
Percent Similarity: 43.36% Conservative: 19
Best Local Similarity: 26.55% Mismatches: 34
Query Match: 13.98% Indels: 30
DB: 10 Gaps: 7

US-09-830-244B-1 (1-95) x US-09-841-132-538 (1-1278)
Qy 2 TrpPheHis-----LysValGlyArg-----LysGlnHisPheLys 13
Db 325 TGGGAACACACTAAACAGCTGGCGATATTCCTCTACTTTTGAAGAGCTATCTTTTGA 384
Qy 14 ValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuValSerLeuLysLys 33
Db 385 GCCTCT-----TCTTCTCAGAAATCCTTATTGAACTCAAGAACCC 426
Qy 34 LysProPheHisLeu-----TyrCysValIleTyrIleProLeuVal 47
Db 427 GAGCCTCAACTATTGGCGATATTAGCCTCTCGGTTTTCGTGTGTCGTCAGAAAT 486
Qy 48 ProLysLeuIleLeuPheLeuAspIleAala---PheIleProLysSerLeuIleSer 66
Db 487 CCT-----TTTCTTTCTCTGGACCTTTTATGCCAAAAACCTATGTGCAA 531
Qy 67 -----GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
Db 532 GGGCAAAACGCTGCTTCTACAAAAAACCTTTATTACTATGACCATGCGCATGTGGAATTA 591
Qy 83 AsnAsnIleArgPheAsnIleLeuSerAsnCysArgThr 95
Db 592 CATTCATAGACTTTCGCATCATTCCTCCACATTTACACA 630

```

```

RESULT 12
US-09-841-132-359
; Sequence 359, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 359
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-841-132-359

Alignment Scores:
Pred. No.: 8.48 Length: 1311
Score: 73.00 Matches: 30
Percent Similarity: 43.36% Conservative: 19
Best Local Similarity: 26.55% Mismatches: 34
Query Match: 13.98% Indels: 30
DB: 10 Gaps: 7

US-09-830-244B-1 (1-95) x US-09-841-132-359 (1-1311)
Qy 2 TrpPheHis-----LysValGlyArg-----LysGlnHisPheLys 13
Db 349 TGGGAACACACTAAACAGCTGGCGATATTCCTCTACTTTTGAAGAGCTATCTTTTGA 408
Qy 14 ValThrPheTrpGluThrAspLeuSerAsnAsnLysThrLeuValSerLeuLysLys 33
Db 409 GCCTCT-----TCTTCTCAGAAATCCTTATTGAACTCAAGAACCC 450
Qy 34 LysProPheHisLeu-----TyrCysValIleTyrIleProLeuVal 47
Db 451 GAGCCTCAACTATTGGCGATATTAGCCTCTCGGTTTTCGTGTGTCGTCAGAAAT 510
Qy 48 ProLysLeuIleLeuPheLeuAspIleAala---PheIleProLysSerLeuIleSer 66
Db 511 CCT-----TTTCTTTCTCTGGACCTTTTATGCCAAAAACCTATGTGCAA 555
Qy 67 -----GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
Db 556 GGGCAAAACGCTGCTTCTACAAAAAACCTTTATTACTATGACCATGCGCATGTGGAATTA 615
Qy 83 AsnAsnIleArgPheAsnIleLeuSerAsnCysArgThr 95
Db 616 CATTCATAGACTTTCGCATCATTCCTCCACATTTACACA 654

RESULT 13
US-09-838-842A-4705
; Sequence 4705, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 22:59:56 ; Search time 83 Seconds
(without alignments)
110.033 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MFHKVGRKHQKVFWEVD.....NHTNHTNNTNIRFNISNRT 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76.5	14.7	2708	2 T09079	probable chloroqui
2	76.5	14.7	2819	2 T09080	probable chloroqui
3	76	14.6	215	2 F84508	probable CCAAT-box
4	76	14.6	580	2 T18439	hypothetical prote
5	76	14.6	600	2 T18446	hypothetical prote
6	75	14.4	163	2 T15985	hypothetical prote
7	73	14.0	426	2 D71552	probable oligopept
8	71	13.6	332	2 T25023	hypothetical prote
9	71	13.6	1274	2 S59405	probable membrane
10	70.5	13.5	567	2 S58750	NADH2 dehydrogenas
11	69.5	13.3	446	2 H90094	hypothetical prote
12	69.5	13.3	1017	2 T18488	hypothetical prote
13	69	13.2	454	2 T32974	protein F37B4.8 (i
14	69	13.2	665	2 E88379	GAF domain protein
15	69	13.2	1245	2 D71613	hypothetical prote
16	68.5	13.1	232	2 E90134	probable UTP-gluc
17	68.5	13.1	499	1 S30007	hypothetical prote
18	68	13.0	204	2 T24668	aspartate-semialde
19	68	13.0	346	1 E64668	oligopeptidase B (
20	68	13.0	683	2 AC0741	hypothetical prote
21	68	13.0	907	2 AE1352	hypothetical prote
22	67.5	12.9	346	2 A70144	membrane-associate
23	67.5	12.9	482	2 F97239	hypothetical prote
24	67	12.8	309	2 T24804	hypothetical prote
25	67	12.8	765	2 S72278	hypothetical prote
26	67	12.8	950	2 S72284	ATP-dependent Clp
27	67	12.8	1028	2 A56038	DNA-directed RNA p
28	67	12.8	1213	2 S16356	DNA-binding protei
29	66.5	12.7	146	2 H72167	ovo protein - frui
30					A35R protein - var

30 66.5 12.7 282 2 A71675
31 66.5 12.7 539 2 B90595
32 66.5 12.7 624 2 T28423
33 66.5 12.7 678 2 S14173
34 66.5 12.7 1238 2 A64596
35 66 12.6 490 2 A46391
36 66 12.6 2150 2 S71629
37 65.5 12.5 306 2 T24402
38 65.5 12.5 463 2 T31570
39 65.5 12.5 2339 2 A45597
40 65 12.5 398 1 R3BYM1
41 65 12.5 618 2 S53384
42 65 12.5 911 2 S28098
43 64.5 12.4 200 2 G97048
44 64.5 12.4 732 2 T23505
45 64 12.3 273 2 A64238

ALIGNMENTS

RESULT 1
T09079

probable chloroquine resistance protein CG2 (strain 7G8) - malaria parasite (Plasm
C:Species: Plasmodium falciparum
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C:Accession: T09079
R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellem, T.E.
Cell 91, 593-603, 1997

A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resi
A:Reference number: Z16556; MUID:98054002; PMID:9393853
A:Accession: T09079

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2708 <SUX>

A:Cross-references: EMBL:AF030692; NID:g2642513; PIDN:AAC47853.1; PID:g2642514

A:Experimental source: strain 7G8; from Brazil

C:Genetics:

A:Gene: cg2

C:Keywords: toxin resistance

Query Match 14.7%; Score 76.5; DB 2; Length 2708;
Best Local Similarity 21.7%; Pred. No. 11;
Matches 25; Conservative 24; Mismatches 37; Indels 29; Gaps 4;
QY 2 WFKVGRKHQKVFWEVDL-----SNKTLVSLKKKPFHL-----YCVIYIP 45
Db 631 WYKIG---YCVIFLFSILKIFCTDDDTLLISRKHVYVLSDSLEKSEYSYKTVIN 686
QY 46 LVPKLIILFDIAFIPKSLISO-----FQNNHYTHNHTNHTNIRFNIIIS 91
Db 687 IHDLLEDFDLHFIFKDELKQDTNGNKYQDMNDHVKNDHTNYQNNIFNLLT 741

RESULT 2
T09080

probable chloroquine resistance protein CG2 (strain HB3) - malaria parasite (Plasm
C:Species: Plasmodium falciparum
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C:Accession: T09080

R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellem, T.E.

Cell 91, 593-603, 1997

A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resi

A:Reference number: Z16556; MUID:98054002; PMID:9393853

A:Accession: T09080

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2819 <SUX>

A:Cross-references: EMBL:AF030693; NID:g2642515; PIDN:AAC47854.1; PID:g2642516

A:Experimental source: strain HB3; from Honduras

C:Genetics:

A:Gene: cg2

C:Keywords: toxin resistance

probable membrane protein YLR436c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein L9753.7
C:Species: *Saccharomyces cerevisiae*
C:Date: 30-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 04-M
C:Accession: S59405
R:Du, Z.

C;Accession: H90094

Tue Apr 15 15:15:24 2003

us-09-830-244b-1.rpr

Page 5

OY 22 SNNKTLVSLKKKPFHLYCVIYIPLVPKLIILFDIAFIPKSLISQFQNNHYTHNH---T 78
Db 1069 SNNYKFFIIRKKKKKKLCYIMKSFHPI----LDEFWNLSCQNEIKNIYYKNLHFVIS 1124
OY 79 NHNTNNIRFNITIS-----NCRT 95
Db 1125 LHNSSIIDFKIINHFIILNKIFENISINCTT 1154

Search completed: April 14, 2003, 23:05:08
Job time : 87 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 22:34:36 ; Search time 84 Seconds
(without alignments)
46.908 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MFHFKVGRKHQKVFWEPTD.....NHTNHTNNTNIRENINISNRT 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
1	68.5	13.1	499	1	UDPGL_YEAST
2	68	13.0	346	1	DHAS_HELPY
3	67	12.8	1028	1	OVO_DROME
4	66.5	12.7	181	1	TBX3_MOUSE
5	66.5	12.7	678	1	SIR1_YEAST
6	66	12.6	490	1	CAR3_DICDI
7	65.5	12.5	396	1	RMAR_YEAST
8	65.5	12.5	2339	1	RPC1_PLAFA
9	65	12.5	618	1	YJK3_YEAST
10	65	12.5	757	1	ECR_LUCU
11	65	12.5	826	1	CRAA_BACUH
12	65	12.5	911	1	STE6_SCHPO
13	64.5	12.4	544	1	PG10_DICDI
14	64	12.3	273	1	ESL3_MYCGE
15	64	12.3	375	1	CAR2_DICDI
16	64	12.3	444	1	PR24_YEAST
17	64	12.3	782	1	ZFY1_MOUSE
18	64	12.3	783	1	ZFY2_MOUSE
19	64	12.3	1177	1	SP97_DICDI
20	63.5	12.2	479	1	YK035W
21	63	12.1	793	1	REGA_DICDI
22	63	12.1	1407	1	CYAA_DICDI
23	62.5	12.0	490	1	PIT_BUCAI
24	62.5	12.0	1570	1	P3K1_DICDI
25	62	11.9	742	1	ZFA_MOUSE
26	62	11.9	969	1	DPOM_NEVIN
27	61.5	11.8	516	1	YK7_YEAST
28	61.5	11.8	567	1	NU2M_HANWI
29	61.5	11.8	1901	1	YCF1_TOBAC
30	61	11.7	237	1	YPV4_METTF
31	61	11.7	317	1	AAC4_DICDI
32	61	11.7	322	1	VF05_VARV
33	61	11.7	448	1	AAC2_DICDI

34	61	11.7	491	1	CAD3_BOVIN
35	61	11.7	658	1	REP_BUCAP
36	61	11.7	666	1	YK2B_HUMAN
37	61	11.7	1427	1	SRB8_YEAST
38	60.5	11.6	467	1	NORA_CHLPN
39	60.5	11.6	686	1	PTRB_ECOLI
40	60	11.5	148	1	Y722_RICPR
41	60	11.5	407	1	RP11_YEAST
42	60	11.5	490	1	CALR_HUMAN
43	60	11.5	554	1	YGS6_YEAST
44	60	11.5	603	1	HNIA_XENLA
45	60	11.5	666	1	YEA7_YEAST

P19535	bos taurus
O31889	buchnera ap
P38401	homo sapien
P25648	saccharomyc
Q92792	chlamydia p
P24555	escherichia
Q92677	ricettisia
P23250	saccharomyc
P30988	homo sapien
P46945	saccharomyc
Q05041	xenopus lae
P40002	saccharomyc

ALIGNMENTS

RESULT 1
UDPG_YEAST
ID UDPG_YEAST STANDARD; PRT; 499 AA.
AC P32861;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Probable UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)
DE (UDP-glucose pyrophosphorylase) (UDPGP).
GN YKL035W OR YKL248.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=93127731; PubMed=1481573;
RA Purnelle B., Skala J., van Dyck L., Goffeau A.;
RT "The sequence of a 12 kb fragment on the left arm of yeast chromosome XI reveals five new open reading frames, including a zinc finger protein and a homolog of the UDP-glucose pyrophosphorylase from potato.";
RT Yeast 8:977-986(1992).
RL
CC -!- FUNCTION: PATHWAYS A CENTRAL ROLE AS A GLUCOSYL DONOR IN CELLULAR METABOLIC PATHWAYS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate -> diphosphate + UDP-glucose.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC UDPGP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X69584; CAA49303.1; -;
CC EMBL; 228035; CAA81872.1; -;
CC PIR; S30007; S30007.
CC SGD; S0001518; YKL035W.
CC InterPro; IPR002618; UDPGP.
CC Pfam; PF01704; UDPGP; 1.
KW Transferase; Kinase; Nucleotidyltransferase; Multigene family.
SQ SEQUENCE 499 AA; 55988 MW; 9B9F4BED885E004D CRC64;
Query Match 13.1%; Score 68.5; DB 1; Length 499;
Best Local Similarity 36.8%; Pred. No. 7.3;
Matches 14; Conservative 8; Mismatches 13; Indels 3; Gaps 1;
OY 52 ILELDIAFIPKSLISQFQNNHYTHNHTNNTNIRNFI 89
DB 291 VRLLEVAQVPKEHIDEFKN---IRKFTNFTNNTNLNL 325

RESULT 2

DHAS_HELPY STANDARD; PRT; 346 AA.

AC 025801; 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (ASA dehydrogenase)

DE (ASADH)

GN ASD OR HP1189.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G., B.A.,

RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,

RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter

RT pylori".

RL Nature 388:539-547(1997).

CC -!- CATALYTIC ACTIVITY: L-aspartate-4-semialdehyde + phosphate +

CC NADP(+) = L-4-aspartyl phosphate + NADPH.

CC -!- PATHWAY: SECOND STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING

CC FROM ASP TO THE CELL WALL PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,

CC TO MET, TO ILE AND TO THR.

CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE ASPARTATE-SEMIALDEHYDE DEHYDROGENASE

CC FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE00625; AAD08235.1; -

DR HSSP; P00353; 1BRM.

DR TIGR; HP1189; -

DR InterPro; IPR000319; Asp_semiald_dh.

DR InterPro; IPR000534; Semiald_dh.

DR Pfam; PF01118; Semialdehyde_dh; 1.

DR Pfam; PF02774; Semialdehyde_dhc; 1.

DR TIGRfam; TIGR01296; asd_B; 1.

DR PROSITE; PS01103; ASD; 1.

KW Oxidoreductase; NADP; Diminopimelate biosynthesis;

KW Lysine biosynthesis; Complete proteome.

FT ACT_SITE 131 131 BY SIMILARITY.

SQ SEQUENCE 346 AA; 38116 MW; EAF10B9108A72330 CRC64;

Query Match 13.0%; Score 68; DB 1; Length 346;

Best Local Similarity 26.3%; Pred. No. 5.7;

Matches 26; Conservative 12; Mismatches 45; Indels 16; Gaps 3;

QY 11 HFVY-----TFWETDLSNNKTVLSKKKPPHLYCVIPLVKKLIIL-----FLDIAT 59

DB 146 HFKIKSVIVSTYQAVSGAGNKGIESLKNELKTALECLEKDPIDLNQVLOAGAFAYPIAT 205

QY 60 IPKSLISQFQNNHYTN-----HTNNTNNTNFRNFIISNC 93

DB 206 NATAHIDTFKENGTYTEELKMLHETHKINGVDFFPISATC 244

RESULT 3

OVO_DROME STANDARD; PRT; 1028 AA.

AC PS1521; Q9X2U4; 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE OVO protein (Shaven baby protein).

GN OVO OR SVB.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RX MEDLINE=95021209; PubMed=7935398;

RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;

RT "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity.";

RL Mol. Cell. Biol. 14:6809-6818(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R;

RX MEDLINE=91293102; PubMed=1712294;

RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;

RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development.";

RL EMBO J. 10:2259-2266(1991).

CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND ACCUMULATES IN NURSE CELLS DURING OÖGENESIS. STORED IN THE EGG, BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U11383; AAB60216.1; -

DR EMBL; X59772; CAB36921.1; ALT_SEQ.

DR HSSP; P07248; 2ADR.

DR TRANSFAC; T00669; -

DR Flybase; FBgn0003028; ovo.

DR InterPro; IPR000822; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 4.

DR PRINTS; PR00048; ZINC_FINGER.

DR SMART; SMD0355; Znf_C2H2; 4.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.

KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;

KW Transcription regulation.

FT DOMAIN 62 66 POLY-ALA.

FT DOMAIN 72 77 POLY-GLY.

FT DOMAIN 80 85 POLY-GLY.

FT DOMAIN 98 108 POLY-GLY.

FT DOMAIN 144 152 POLY-HIS.

FT DOMAIN 153 159 POLY-ASN.

FT DOMAIN 336 339 POLY-GLN.

FT DOMAIN 347 353 POLY-GLN.

FT DOMAIN 357 361 POLY-GLN.

FT DOMAIN 410 414 POLY-GLN.

FT DOMAIN 418 422 POLY-GLN.

FT DOMAIN 426 432 POLY-GLN.

FT DOMAIN 445 453 POLY-GLN.

DR EMBL; V00705; CAA24077.1; -
 DR PIR; A02751; R3BYM1.
 DR SGD; S0007275; VARI.
 KW Ribosomal protein; Mitochondrion.
 SQ SEQUENCE 396 AA; 46881 MW; 62A68A24740BCBEO CRC64;
 Query Match 12.5%; Score 65.5; DB 1; Length 396;
 Best Local Similarity 23.18; Pred. No. 12;
 Matches 27; Conservative 21; Mismatches 36; Indels 33; Gaps 6;
 QY 4 HKVG-----RKQHF-KVTWETDLN-NKTLVSLKKKPFHLYCVIYIPLVPLKLIIF 54
 DB 72 HRIGPAGNINKLQHLNNMNWNTQIYNKNMEIMTMDKLINKLLY-----KMTLK 126
 QY 55 LDIAFIPKSLISQFQNNH-----YTHNHTNHTNIRF-----NIISN 92
 DB 127 LNNMINKITMSKTINQHSLSKLNKIKFYNNNDINNNDNNNNNNNNNNNNNNNNNNNN 183
 RESULT 8
 RCL_PLAFA STANDARD; PRT; 2339 AA.
 ID RCL_PLAFA
 AC P27625;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92018020; PubMed=1656254;
 RA Li W.B., Bizik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;
 RT "Characterization of the gene encoding the largest subunit of
 Plasmodium falciparum RNA polymerase III";
 RL Mol. Biochem. Parasitol. 46:229-240(1991).
 CC -|- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 (RNA)(N).
 CC -|- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
 SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE
 III.
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 III FOR 5S AND TRNA GENES.
 CC -|- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M73770; AAA29729.1; -
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR002879; RNA_pol_A2.
 DR Pfam; PF00623; RNA_pol_A; 1.
 DR Pfam; PF01854; RNA_pol_A2; 2.
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Zinc-finger; Nuclear protein.
 FT ZN_FING 88 101 C3H-TYPE (POTENTIAL).
 SQ SEQUENCE 2339 AA; 272829 MW; EDD899363086FD48 CRC64;
 Query Match 12.5%; Score 65.5; DB 1; Length 2339;
 Best Local Similarity 20.2%; Pred. No. 74;
 Matches 20; Conservative 15; Mismatches 23; Indels 41; Gaps 4;

QY 4 HKVGRKQHFVYFWETDLNN-----KTLVSL--KKKKPFHLYCVIYIPLVP 48
 DB 1460 NNIYRNEKKKLNKRWKTKMDNDNYWSSDDSIIAKKIKIKKKEKKRYH----- 1508
 QY 49 KLILFLDIAFIPKSLISQFQNNHY---THNHTNHTN 84
 DB 1509 -----PKEKENFDRNNYKMITDNNNDNNNN 1535
 RESULT 9
 YJK3_YEAST STANDARD; PRT; 618 AA.
 ID YJK3_YEAST
 AC P42950;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative 70.4 kDa transcriptional regulatory protein in IME2-MEF2
 DE intergenic region.
 DE YJL103C OR J0824.
 GN YJL103C
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / FY1679;
 RX MEDLINE=96090136; PubMed=7483851;
 RA Rasmussen S.W.;
 RT "A 37.5 kb region of yeast chromosome X includes the SMEL, MEF2, GSH1
 and CSD3 genes, a TCP-1-related gene, an open reading frame similar
 to the DAL80 gene, and a trna(Arg).";
 RL Yeast 11:873-883(1995).
 CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -|- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
 CLUSTER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X65021; CAA59391.1; -
 DR EMBL; Z49378; CAA89398.1; -
 DR SGD; S0003639; YJL103C.
 DR InterPro; IPR001138; Fungi_Trn.
 DR Pfam; PF00172; Zn_clus; 1.
 DR SMART; SM00066; GAL4; 1.
 DR PROSITE; PS00463; ZN2_Cy6_FUNGAL_1; 1.
 DR PROSITE; PS00448; ZN2_Cy6_FUNGAL_2; 1.
 KW Hypothetical protein; Transcription regulation; DNA-binding;
 KW Nuclear protein; Zinc; Metal-binding; Sporulation; Transmembrane.
 FT DNA_BIND 20 48 ZN(2)-CYS(6), FUNGAL-TYPE.
 SQ SEQUENCE 618 AA; 70381 MW; 200F6AEE06582E38 CRC64;
 Query Match 12.5%; Score 65; DB 1; Length 618;
 Best Local Similarity 32.9%; Pred. No. 21;
 Matches 23; Conservative 12; Mismatches 21; Indels 14; Gaps 4;
 QY 20 DLSNKTLSLKKKPFHLYCVIYIPLVPLKLIIFLDIAFIPKSLISQFQNNHTN 79
 DB 373 DLYDNKCLI-----KP-HNYKLAYKLTTLTKKFLKFLKFLKFLKFLKFLKFLKFLK 422
 QY 80 HNTNIRFNI 89
 DB 423 ----NLYDL 428
 RESULT 10
 ECR_LUCCU

```

ID AC ECR_LUCCO STANDARD; PRT; 757 AA.
DT OI8531;
DT DT 15-DEC-1998 (Rel. 37, Created)
DT DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-eecdysone
GN GN receptor) (20E receptor).
OS OS BCR OR NRH1.
OC OC Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OX X Muscomorpha; Gasteroidea; Calliphoridae; Lucillia.
RN RN [1]
RP RP SEQUENCE FROM N.A.
RX RX MEDLINE=97449774; PubMed=9304790;
RA RA Hannan G.N., Hill R.J.;
RT RT "Cloning and characterization of LcEcR: a functional ecdysone
RL RL receptor from the sheep blowfly Lucilla cuprina.";
RC RC Insect Biochem. Mol. Biol. 27:479-488(1997).
CC CC -! FUNCTION: RECEPTOR FOR ECYDSONE. BINDS TO ECYDSONE RESPONSE
CC CC ELEMENTS (ECRES) (BY SIMILARITY).
CC CC -- SUBCELLULAR LOCATION: Nuclear.
CC CC -! SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC CC NRL SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR DR EMBL; U75355; ABAB1130.1; -.
DR DR HSSP; P20393; IAGB.
DR DR InterPro; IPRO00536; Hormone_rec_lig.
DR DR InterPro; IPRO01723; Stdhrmn_receptor.
DR DR InterPro; IPRO01628; Znf_C4steroid.
DR DR Pfam; PF00104; hormone_rec; 1.
DR DR Pfam; PF00105; zf-C4; 1.
DR DR PRINTS; PR00398; STRDHOMONER.
DR DR PRINTS; PR00047; STROIDFINGER.
DR DR PRODOM; PD000035; znf_C4steroid; 1.
DR DR SMART; SM00430; HOL1; 1.
DR DR SMART; SM00399; znf_C4; 1.
DR DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
FT FT Zinc-finger.
FT FT DOMAIN 1 300 MODULATING (POTENTIAL).
FT FT DNAS_BIND 301 366 NUCLEAR RECEPTOR-TYPE.
FT FT ZNF_FING 301 321 C4-TYPE.
FT FT ZN_FING 337 361 C4-TYPE.
FT FT DOMAIN 454 674 HORMONE-BINDING (POTENTIAL).
SQ SQ SEQUENCE 757 AA; 83075 MW; C151145ZED3D359 CRC64;

Query Match 12.5%; Score 65; DB 1; Length 757;
Best Local Similarity 40.7%; Pred.No. 26;
Matches 11; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY QY 66 SOFONNHVTHNTHTNNTRFNLIIS 92
Db Db 152 SOVVHHNSHNHLHHSNSNHSSSH 178
||| :|||| |:::|| ||| ::| |
RESULT 11
CRAA_BACUH STANDARD; PRT; 826 AA.
ID ID CRAA_BACUH
AC AC Q9S597;
DT DT 16-OCT-2001 (Rel. 40, Created)
DT DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DE Pesticidal crystal protein cry27Aa(Insecticidal delta-endotoxin
```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gables S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hillbert H.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Pohl T.M.,
RA Borzym K., Langer I., Beck A., Lehrach H., Wambutt R., Pohl T.M.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN IS ESSENTIAL FOR MATING.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53254; CAA37345.1; -
CC EMBL; AL049559; CAB40184.1; -
CC EMBL; AL031966; CAA21435.1; -
CC PIR; S28098; S28098.
CC InterPro; IPR000851; RasGEF.
CC InterPro; IPR001895; RasGRF_CDC25.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEF; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00147; RasGEF; 1.
CC SMART; SM00229; RasGEF; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00720; GDS_CDC25; 1.
CC PROSITE; PS00002; SH3; 1.
KW Guanine-nucleotide releasing factor; SH3 domain.
FT DOMAIN 1 60
SQ SEQUENCE 911 AA; 105185 MW; 5F8C12D20C4B753F CRC64;

Query Match 12.5%; Score 65; DB 1; Length 911;
Best Local Similarity 25.2%; Pred. No. 32;
Matches 28; Conservative 10; Mismatches 23; Indels 50; Gaps 5;

Qy 23 NKTIVLSL-----KKKKPFHY-----CVIYIPLVPLKLIILFDIAFIPK----- 62
Db 776 NSKTLASFELLNNLTARKNFKNFYRDLNCLNCCV--LPCVPFLGVYFDLTFLKTCNKDNQ 833
Qy 63 -----SLISQFNHHYHNHN-----HNTNNI 85
Db 834 NAINPDKRTKVTIRLNILKFKQSVGYMFPNINEVQELLNEVISRERNNNI 884

RESULT 13
GP10_DICDI
ID GP10_DICDI STANDARD; PRT; 544 AA.

AC Q05885;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Glycoprotein GP100 precursor (P29F8).
GN GPPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=941124612; PubMed=8294503;
RA Barth A., Mueller-Taubenberger A., Taranto P., Gerisch G.;
RT "Replacement of the phospholipid-anchor in the contact site A
RT glycoprotein of D. discoideum by a transmembrane region does not
RT impede cell adhesion but reduces residence time on the cell
RT surface.";
RL J. Cell Biol. 124:205-215(1994).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: RESTRICTED TO THE AGGREGATION STAGE OF
CC DEVELOPMENT IN D.DISCOIDEUM. NO DETECTABLE ACTIVITY IN CELL
CC ADHESION.
CC -!- FTM: N- AND O-GLYCOSYLATED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L04286; AAC37369.1; -
CC DictyDb; D02035; gppa.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 544
FT DOMAIN 20 489
FT TRANSMEM 490 510
FT DOMAIN 511 544
FT DOMAIN 117 208
FT CARBOHYD 80 80
FT CARBOHYD 224 224
FT CARBOHYD 308 308
FT CARBOHYD 332 332
FT CARBOHYD 366 366
FT CARBOHYD 380 380
FT CARBOHYD 410 410
FT CARBOHYD 422 422
FT CARBOHYD 478 478
SQ SEQUENCE 544 AA; 59156 MW; 9B2C2CDE7C7D0AA0 CRC64;

Query Match 12.4%; Score 64.5; DB 1; Length 544;
Best Local Similarity 26.9%; Pred. No. 21;
Matches 21; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

Qy 18 ETDLSNNKTLVSLKKKPFHYCVIYIPLVPLKLIILFDIAFIPKLSIQFNHHYHNH 77
Db 472 DSKSNNNSGGGKNNK-----LYLLIPLVPLVLIIVAAVAIFI-KTRVSGNSGSKVKNKN 526
Qy 78 TNHNTNNRNFNIISNCR 95
Db 527 NKKDSINVPFQMLDEITT 544

RESULT 14
ESL3_MYCGE
ID ESL3_MYCGE STANDARD; PRT; 273 AA.
AC Q49421; Q49335;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative esterase/lipase 3 (EC 3.1.-.-).

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 22:53:41 ; Search time 110 Seconds
(without alignments)
177.950 Million cell updates/sec

Title: US-09-830-244B-1

Perfect score: 522

Sequence: 1 MFHFVKVRQHKFKVTFWYD.....NHTNHTNIRENIISNCR 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	18.1	512	8 Q9XMS7	Q9XMS7 tetrahymena
2	84	16.1	135	15 Q90814	Q90814 human immun
3	76.5	14.7	2708	5 Q15791	Q15791 plasmodium
4	76.5	14.7	2742	5 Q15801	Q15801 plasmodium
5	76.5	14.7	2819	5 Q15792	Q15792 plasmodium
6	76	14.6	215	10 Q9SIT9	Q9SIT9 arabidopsis
7	76	14.6	573	5 Q77334	Q77334 plasmodium
8	76	14.6	1672	5 Q8T2M6	Q8T2M6 dictyosteli
9	75	14.4	163	5 Q10054	Q10054 caenorhabdi
10	74.5	14.3	721	5 Q8T218	Q8T218 dictyosteli
11	74.5	14.3	2344	5 Q8SSV9	Q8SSV9 dictyosteli
12	74	14.2	141	12 Q9UF67	Q9UF67 vaccinia vi
13	73	14.0	426	16 Q84141	Q84141 chlamydia t
14	73	14.0	1317	5 Q8T4K2	Q8T4K2 dictyosteli
15	72.5	13.9	616	5 Q8TID3	Q8TID3 dictyosteli
16	72.5	13.9	2567	5 Q9U0J6	Q9U0J6 plasmodium

17	71.5	13.7	207	10 Q9AVX6	Q9AVX6 guillardia
18	71.5	13.7	518	8 Q95029	Q95029 tetrahymena
19	71	13.6	215	2 Q93UA8	Q93UA8 carsonella
20	71	13.6	332	5 Q9XUN3	Q9XUN3 caenorhabdi
21	71	13.6	1146	3 Q77382	Q77382 plasmodium
22	71	13.6	1274	3 Q66673	Q66673 saccharomyc
23	70.5	13.5	1264	5 Q8WPV5	Q8WPV5 schistosoma
24	70.5	13.5	3432	5 Q95PH9	Q95PH9 dictyosteli
25	70	13.4	134	12 Q9EMW6	Q9EMW6 amsacta moo
26	69.5	13.3	154	15 Q9E8H5	Q9E8H5 human immun
27	69.5	13.3	343	12 Q9LEW7	Q9LEW7 cydia pomon
28	69.5	13.3	445	10 Q98RQ9	Q98RQ9 guillardia
29	69.5	13.3	548	5 Q8TIS2	Q8TIS2 dictyosteli
30	69.5	13.3	1017	5 Q77371	Q77371 plasmodium
31	69.5	13.3	1384	12 Q9EMR9	Q9EMR9 amsacta moo
32	69.5	13.3	2067	5 Q9U3U8	Q9U3U8 plasmodium
33	69.5	13.3	3417	5 Q8T116	Q8T116 dictyosteli
34	69	13.2	207	8 Q9G8W1	Q9G8W1 rhodomonas
35	69	13.2	454	5 Q45096	Q45096 caenorhabdi
36	69	13.2	605	5 Q45170	Q45170 caenorhabdi
37	69	13.2	1245	5 Q96195	Q96195 plasmodium
38	69	13.2	1281	5 Q15802	Q15802 plasmodium
39	68.5	13.1	215	2 Q93099	Q93099 carsonella
40	68.5	13.1	232	10 Q9S16	Q9S16 guillardia
41	68.5	13.1	502	3 Q9P966	Q9P966 candida gla
42	68.5	13.1	643	5 Q8ST22	Q8ST22 dictyosteli
43	68.5	13.1	743	5 Q9NGX0	Q9NGX0 dictyosteli
44	68.5	13.1	1560	5 Q9GRX5	Q9GRX5 dictyosteli
45	68.5	13.1	1561	5 Q9U987	Q9U987 dictyosteli

ALIGNMENTS

RESULT 1

Q9XMS7	Q9XMS7	PRELIMINARY;	PRT;	512 AA.
ID	Q9XMS7			
AC	Q9XMS7			
DT	01-NOV-1999 (TREMREL. 12, Created)			
DT	01-DEC-2001 (TREMREL. 12, Last sequence update)			
DT	01-DEC-2001 (TREMREL. 19, Last annotation update)			
DE	Haem lyase.			
GN	YEJR.			
OS	Tetrahymena pyriformis.			
OG	Mitochondrion.			
OC	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;			
OC	Tetrahymenina; Tetrahymena.			
OX	NCBI_TaxID=5908;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-20181866; PubMed-10715208;			
RA	Edqvist J., Burger G., Gray M.W.;			
RT	"Expression of mitochondrial protein-coding genes in Tetrahymena pyriformis.";			
RL	J. Mol. Biol. 297:381-393(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-20181865; PubMed-10715207;			
RA	Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,			
RA	Lang B.F., Gray M.W.;			
RT	"Complete sequence of the mitochondrial genome of Tetrahymena pyriformis and comparison with Paramaecium aurelia mitochondrial DNA.";			
RL	J. Mol. Biol. 297:365-380(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Burger G.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF160864; A041936.1;			
DR	InterPro; IPR003567; Cyt_c_biol.			
DR	PRINTS; PR01410; CCBIOGENESIS.			
KW	Lyase; Mitochondrion.			
SQ	SEQUENCE. 512 AA; 62577 MW; 6A4EF0808AC26366 CRC64;			

```

Query Match      18.1%; Score 94.5; DB 8; Length 512;
Best Local Similarity 33.7%; Pred. No. 0, 029;
Matches 28; Conservative 16; Mismatches 30; Indels 9; Gaps 4;

QY 10 QHFKVTEWETDLSNNKTLVSLAKKKPFHLHYCVIYIPLVPKLIIL-FLDIAFIPK---SLI 65
   :||| |||:::|||::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 344 KHKWSNFWFSLLNNEFTIIS-----NYKIYWIRIYKTHTLIIILMFISYIFLNKLINLV 399
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 SQFQNHHYTHNHT-NHNTNNIRF 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 QQINNYYIVLSNTIOGSTNNILF 422
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
CS0814 PRELIMINARY; PRT; 135 AA.
ID O90814 AC O90814;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 2.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11709;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A.
RX MEDLINE=97255645; PubMed=9100992;
RA Xiang Z., Ariyoshi K., Wilkins A., Dias F., Whittle H., Breuer J.;
RT "HIV type 2 pathogenicity is not related to subtype in rural guinea
RL bissau.";
RT AIDS Res. Hum. Retroviruses 13:501-505(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A.
RX MEDLINE=98406190; PubMed=9733826;
RA Grassly N., Xiang Z., Ariyoshi K., Aaby P., Jensen H., Dias F.,
RT Van der Loeff, Whittle H., Breuer J.;
RT "Mortality among human immunodeficiency virus type 2-positive
RT villagers in rural guinea-bissau is correlated with viral genotype.";
RT J. Virol. 72:7895-7899(1998).
DR EMBL: AJ011256; CAA09567.1; -.
DR InterPro: IPR000777; GP120.
DR PFam: PF00516; GP120; 1.
DR AIDS; Coat protein; Glycoprotein.
KW NON_TER 1
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 15598 MW; 2A732E706BBE663B CRC64;

Query Match      16.1%; Score 84; DB 15; Length 135;
Best Local Similarity 25.9%; Pred. No. 0, 12;
Matches 29; Conservative 15; Mismatches 32; Indels 36; Gaps 5;

QY 16 FWETDLSNNKTLVSLAKKKPFHLHC-----VVIPLVPKLIILFLDIAFIPKSLISQF 68
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 YWHS--KNKNTIISLNKYNLHLCKRPNKVTPVTILMSGLVFHSQPTNKPROAQCF 72
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 QN-----NHYTHNHTNHTNNIRFN-----IISNCR 94
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 73 KGQKKAMLEVKQTIVKHKPYKETN-DTNNIHFTAPEKGSNPKVAYMTNCR 123
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
O15791 PRELIMINARY; PRT; 2708 AA.
ID O15791 AC O15791;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
CG2.
GN CG2.
OS Plasmodium falciparum (isolate 7G8).

```

```

Db 643 WYKIG-----YCVFFLSILKIFCTCDDTTLLSRKHVYLLSDSLEKSEYKGTGVN 698
Qy 46 LVPKLIILFLDIAPKSLISQ-----FQNNHYTHNTHNTNINIRFNIS 91
Db 699 IHDLLDYFFDLHFIDKELACQDTNGNIKFQDMNDHVKNDDHTNYQNNIFNLT 753

RESULT 5
O15792 ID O15792 PRELIMINARY; PRT; 2819 AA.
AC O15792;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN CG2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB3;
RX MEDLINE=98054002; PubMed=9393853;
RA Su X., Kirman L.A., Fujioka H., Welles T.E.;
RT "Complex polymorphisms in an approximately kDa protein are linked to
RT chloroquine-resistant P. falciparum in Southeast Asia and Africa.";
RL Cell 91:593-603(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HB3;
RX MEDLINE=20446201; PubMed=10989140;
RA Fidock D.A., Nomura T., Cooper R.A., Su X.-z., Talley A.K.,
RA Welles T.E.;
RT "Allelic modifications of the cg2 and cgl genes do not alter the
RT chloroquine response of drug-resistant Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 110:1-10(2000).
DR EMBL; AF030693; AAC47854.1;
SQ SEQUENCE 2819 AA; 335122 MW; FBF3CCCC0AC788D9 CRC64;

Query Match 14.78; Score 76.5; DB 5; Length 2819;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 25; Conservative 24; Mismatches 37; Indels 29; Gaps 4;

Qy 2 WFKVGRKHQKVFETWETDL-----SNKTLVSLKPKPHL-----YCVIYIP 45
Db 637 WYKIG-----YCVFFLSILKIFCTCDDTTLLSRKHVYLLSDSLEKSEYKGTGVN 692
Qy 46 LVPKLIILFLDIAPKSLISQ-----FQNNHYTHNTHNTNINIRFNIS 91
Db 693 IHDLLDYFFDLHFIDKELACQDTNGNIKFQDMNDHVKNDDHTNYQNNIFNLT 747

RESULT 6
Q9SIT9 ID Q9SIT9 PRELIMINARY; PRT; 215 AA.
AC Q9SIT9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative CCAAT-box-binding transcription factor.
GN AT2G13570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
```

```

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007063; AAD22680.1;
DR HSSP; P48781; 1b67.
DR InterPro; IPR003956; CBFA_NFYB.
DR InterPro; IPR003958; CBFA_NFYB_domain.
DR InterPro; IPR003957; CBFA_NFYB_topis.
DR InterPro; IPR004822; Histone_core.
DR Pfam; PF00808; CBFD_NFYB_HMF. 1.
DR PRINTS; PR00615; CCATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB. 1.
SQ SEQUENCE 215 AA; 24619 MW; 0615B2B7A097DD25 CRC64;

Query Match 14.6%; Score 76; DB 10; Length 215;
Best Local Similarity 27.0%; Pred. No. 1.3;
Matches 24; Conservative 14; Mismatches 27; Indels 24; Gaps 3;

Qy 5 KVGKQHKVFETWETDLNNKTLVSLKPKPHLYCVIYI-----PLVPKLIILFLDIA 58
Db 145 QIQQNHNYQFOEQDQNN-----NMCTSYISHHHPFLPVDHPQPPNIA 192
Qy 59 FIPKSLISQFQNNHYTHNTHNTNINIRF 87
Db 193 FSPKSLQKQFPQOH-----DNNIDSIH 215

RESULT 7
O77334 ID O77334 PRELIMINARY; PRT; 573 AA.
AC O77334;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative dual-specificity protein phosphatase.
GN MAL3p3.11.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajadream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538(1999).
DR EMBL; Z98547; CAB1119.2;
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSPC; 1.
DR SMART; SM00012; PTPC_DSPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00054; TYR_PHOSPHATASE_DUAL; 2.
SQ SEQUENCE 573 AA; 67573 MW; 97E22AUAFA37D3334 CRC64;
```

RA	Smdalon N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.; "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans"; Nature 368:32-38(1994).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BRISTOL N2;
RA	Anderson K.;
RT	"The sequence of C. elegans cosmid F09C12.";
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BRISTOL N2;
RA	Waterston R.;
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR	EMBL; U28929; AAA68347.1; -;
SQ	SEQUENCE 163 AA; 18602 MW; 2305A1F3E7290CB5 CRC64;
Query Match 14.4%; Score 75; DB 5; Length 163;	
Best Local Similarity 25.2%; Pred. No. 1.3;	
Matches 26; Conservative 20; Mismatches 37; Indels 20; Gaps	
QY	1 MFPHKVGKRQHFKFWETDLSNNKTLYSLKKPKHLYCVIYIPVLPKLIIFLDLTAFI 60 : : : : : : : : : : : : : :
Db	63 INFHKNGRR-----FENNDAIVGKIFRSSEFLASTVLITLLIGVF-IAYR 111 : : : : : : : : : : : : : :
QY	61 PKSL-----ISOFQNNHYTHNTNNTNRFNINSCRT 95 ::: :
Db	112 KRNVLVYCGRKAAFGNDYATFYI-CNISGYLFSFLNCIT 153 : : : : : : : : : : : : : : : : : : : :
RESULT 10	
Q87218	PRELIMINARY; PRT; 721 AA.
ID	Q87218
AC	Q87218
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Hypothetical 83.4 kDa protein.
OS	Dictyostellium discoideum (Slime mold).
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX	NCBI_Taxid=44689;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-A4;
RA	Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P., Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A. "Sequence and Analysis of Chromosome 2 of Dictyostellium."; Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AC116030; AAL92987.1; -;
KW	Hypothetical protein.
SQ	SEQUENCE 721 AA; 83384 MW; AY8C83BE1FE5BC1 CRC64;
Query Match 14.3%; Score 74.5; DB 5; Length 721;	
Best Local Similarity 29.1%; Pred. No. 6.1;	
Matches 30; Conservative 18; Mismatches 30; Indels 25; Gaps	
QY	11 HFK---VTFEW----TDL---SNNKTLVSLKKKPHELYC-----VIYIPVPKLIIFLD 56 :
Db	149 HFKGVMLSNEKFIIDLFFNNNNNNNNNNPLEKFLQLYNSPNEILNPTMD-----N 201 :
QY	57 IAFTPKLSIQF--QNHHY--THNHTNNTNRFNINSCRT 95 : : : : : : : : : : : : : : : : : : : :
Db	202 ISFTSHLLFDLYNNNNNNYNINSNNNNNNNNNNNNNNNNNST 244 : : : : : : : : : : : : : : : : : : : :
RESULT 11	
Q85SV9	PRELIMINARY; PRT; 2344 AA.
ID	Q85SV9
AC	Q85SV9;

```

DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DE Putative transferase.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lemann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunngal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequenced and Analysis of Chromosome 2 of Dictyostellium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC115598; AAL93230.1; -
SQ SEQUENCE 2344 AA; 266642 MW; A897406B0A4634 CRC64;

Query Match 14.3%; Score 74.5; DB 5; Length 2344;
Best Local Similarity 27.5%; Pred. No. 18;
Matches 28; Conservative 15; Mismatches 28; Indels 31; Gaps 6;

QY 2 WFHKVGRKHQKVFETDLSNNKTLVSLKK-KP-----FHLVCYIYIPLV 47
Db 1697 WFTKC-----HFETI-----NKTIRAFKQNKPTSPDVLHDLIAEFQYRLISLSLS 1744
QY 48 ---PKLIILFLDIATFKPKLSIQFQNN--HYTHNHTNNTNN 84
Db 1745 SVNPSLALFRPTITQPGFDLNYFNINHHHHHHHHHGN 1786

RESULT 12
QY 09JF67 PRELIMINARY; PRT; 141 AA.
AC Q9JF67;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE T4A0R.
OS Vaccinia virus (strain Tian Tan).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10253;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIAN TAN;
RA Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;
RT "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF095689; AAF34036.1; -
DR InterPro: IPR000504; RNA_rec_mot.
DR PROSITE: PS00030; RRM_RNP.1; UNKNOWN.1.
SQ SEQUENCE 141 AA; 16504 MW; 3B0DF51BDF9A0231 CRC64;

Query Match 14.2%; Score 74; DB 12; Length 141;
Best Local Similarity 18.5%; Pred. No. 1.5;
Matches 23; Conservative 27; Mismatches 30; Indels 44; Gaps 4;

QY 13 KVTFWETDLSNNKTLVSLKKK-----KPFHLYCYIYI-----PLVP----- 48
Db 13 KTSFYNCNDSTREKIKIKHKGMSFVYKPKHSTVYKGLSGGGIYHDDLVLGKVTINDL 72
QY 49 KLIILFLDIATFKPKLSIQFQNN--HYTHNHTNNTNN 84
Db 73 KMLFYMDLSYHGVTSSGVYKLGSSIDRLSLNRTIVTKVNNYNNYNNYNNYNNYNN 132
QY 89 IISN 92
Db 133 NYNN 136

RESULT 13

```

```

084141
ID 084141 PRELIMINARY; PRT; 426 AA.
AC 084141;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DE Oligopeptide binding protein.
OS OPPA_1 OR Ctl139.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.;
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AE001287; AAC67730.1; -
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
KW Complete proteome.
SQ SEQUENCE 426 AA; 48271 MW; 8C1FF71717980235 CRC64;

Query Match 14.0%; Score 73; DB 16; Length 426;
Best Local Similarity 26.5%; Pred. No. 5.4;
Matches 30; Conservative 19; Mismatches 34; Indels 30; Gaps 7;

QY 2 WFH--KVGR-----KQHEKVFETDLSNNKTLVSLKKKPFHL-----YCVIYIPLV 47
Db 109 WEHTKAGRYSLLEKLSFRAS-----SSSELLHELKEPEQLAILAISPFAVYRPN 162
QY 48 PKLIILFLDIA-FIPKSLIS----QFQNNHYTHNHTNNTNNIRNITSNCRT 95
Db 163 P-----FLSSGPFMPKTYVQGTVLVQKNPYDYDHAHVELHSIDFRIIPNIT 210

RESULT 14
QY 08T4K2 PRELIMINARY; PRT; 1317 AA.
AC 08T4K2;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DE ABC transporter ABCC14.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporter of Dictyostellium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY082376; AAL89812.1; -
SQ SEQUENCE 1317 AA; 150238 MW; 1768BBFD23B24D1A CRC64;

Query Match 14.0%; Score 73; DB 5; Length 1317;
Best Local Similarity 23.8%; Pred. No. 16;
Matches 31; Conservative 16; Mismatches 39; Indels 44; Gaps 5;

QY 8 RKQHFK-----VTFWETD-----LSNNKTLVSLKKKPFHL-----YCVIY 43
Db 330 RKQELKFLFLVLFWIFDHHMIETNATLVLVSTFATVLTGNTMSLDVTFTAMTIFANLK 389
QY 44 IPLVPKLIILFLDIATFKPKLSIQFQNN--HYTHNHTNNTNNIRN--- 88
Db 390 LPLIYLPEDIYKAIGLMPKVKIQNFKLSKESLKSRENQINFNINNNNNNNNNNNNN 449
QY 89 -----IISNC 93

```

```
Db 450 DDNDIIENC 459
|| ||
RESULT 15
Q8TID3
ID Q8TID3 PRELIMINARY; PRT; 616 AA.
AC Q8TID3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 71.6 kDa protein
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116960; AAM08465.1; -.
KW Hypothetical protein.
SQ SEQUENCE 616 AA; 71626 MW; 8D92837ACBF3E045 CRC64;

Query Match 13.9%; Score 72.5; DB 5; Length 616;
Best Local Similarity 35.0%; Pred. No. 8.7;
Matches 21; Conservative 9; Mismatches 27; Indels 3; Gaps 1;

QY 33 KKPEHLYCVIYPLVPKLIILFLDIAFTPKSLISQFONHHYTHNHTNNTNIRFTISN 92
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 285 KSWHIYIVGLIESFPLISLISLGDIS---KENINNNNNNNNNNNNNNNNNNNNN 341

Search completed: April 14, 2003, 23:03:32
Job time : 114 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 14, 2003, 23:06:51 ; Search time 147.7 Seconds
(without alignments)
1043.807 Million cell updates/sec

Title: US-09-830-244B-1
Perfect score: 522
Sequence: 1 MWFKVGRKQHFVTFWTD.....NHTNHTNIRNFNIISNCR 95

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/Cgn2_1/USFPO.spool/US09830244/runat_10042003_090016.27874/app.query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi-LIST=45
-DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830244.CGN_1.1.2874 @runat_10042003_090016.27874 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_esthu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	349	66.9	306	10	BE163886	BE163886 OV3-HT046
2	321	61.5	951	12	BE168510	BE168510 602339663
c 3	294	56.3	329	14	BQ309679	BQ309679 RC6-BT070
4	201	38.5	615	12	BF678862	BF678862 602153173
5	180.5	34.6	286	9	AA376683	AA376683 EST89136
6	178.5	34.2	304	9	AA376720	AA376720 EST89178
7	178.5	34.2	341	9	AA376695	AA376695 EST89151
8	178.5	34.2	426	9	AA376680	AA376680 EST89133
9	177	33.9	321	12	BG205925	BG205925 RST25360
10	176	33.7	335	9	AA376714	AA376714 EST89171
11	176	33.7	470	12	BG200817	BG200817 RST20025
12	174.5	33.4	517	12	BE188342	BE188342 RST7360 A
13	173.5	33.2	519	10	AW951480	AW951480 EST363550
14	171.5	32.9	457	9	AA376682	AA376682 EST89135
c 15	164	31.4	404	9	AA376713	AA376713 EST89170
c 16	156	29.9	405	9	AA376719	AA376719 EST89177
c 17	150	28.7	375	9	AA376681	AA376681 EST89134
18	148	28.4	714	10	BE253236	BE253236 601116852
c 19	145	27.8	410	14	T29650	T29650 EST89132 Hu
c 20	144.5	27.7	514	9	AA195115	AA195115 ZR35B03.F
c 21	144	27.6	334	9	AA744925	AA744925 np72e2.s
c 22	144	27.6	373	10	BE464523	BE464523 hv84h04.x
c 23	144	27.6	440	9	AI421621	AI421621 tf38g10.x
c 24	144	27.6	449	9	AA747799	AA747799 nx86a06.s
c 25	144	27.6	458	9	AI150564	AI150564 qf36e06.x
c 26	144	27.6	461	9	AI377872	AI377872 te63h11.x
c 27	144	27.6	477	13	BM663533	BM663533 UT-E-CO-
28	144	27.6	495	14	BQ427792	BQ427792 AGENCOURT
c 29	144	27.6	496	10	AW444891	AW444891 UI-H-B13-
c 30	144	27.6	519	14	BM966861	BM966861 IJ27d11.y
c 31	144	27.6	559	14	BM939624	BM939624 IJ27d11.y
c 32	143.5	27.5	373	12	BF477953	BF477953 7-08A06.x
c 33	143.5	27.5	394	12	BF115130	BF115130 h7405.x
c 34	143.5	27.5	427	12	BF477850	BF477850 7-06A06.x
c 35	143.5	27.5	428	10	AW613151	AW613151 bh42406.x
c 36	143.5	27.5	451	10	AW189807	AW189807 x109C07.x
c 37	143.5	27.5	469	10	AW293206	AW293206 UI-H-B12-
c 38	143.5	27.5	513	9	AA195114	AA195114 z35a04.f
c 39	143.5	27.5	613	10	AV727718	AV727718 AV727718
c 40	143	27.4	321	10	AW001596	AW001596 wr99g09.x
c 41	143	27.4	440	10	AW468605	AW468605 he42b05.x
c 42	143	27.4	619	9	AA961420	AA961420 OF53a12.s
c 43	142.5	27.3	521	10	BE465101	BE465101 hv75b07.x
c 44	142	27.2	374	10	BE247773	BE247773 TCBA2P253
c 45	141	27.0	373	9	AI692383	AI692383 wd63g06.x

ALIGNMENTS

RESULT 1
LOCUS BE163886/C 306 bp mRNA linear EST 21-JUN-2000
DEFINITION QV3-HT0461-230200-103-h02 HT0461 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE163886
VERSION BE163886.1 GI:8626607
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 306)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6at2-RC6-BT0709-310300-021-F02&t3=2000-03-31&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 22

High quality sequence stop: 57.

FEATURES

source

1. 329
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0709"
/dev_stage="Adult"

/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 106 a 53 c 60 g 110 t

ORIGIN

Alignment Scores:

Pred. No.: 7 93e-26 Length: 329
Score: 294.00 Matches: 60
Percent Similarity: 96.83% Conservative: 1
Best Local Similarity: 95.24% Mismatches: 0
Query Match: 56.32% Indels: 2
DB: 14 Gaps: 0

US-09-830-244B-1 (1-95) x BQ309679 (1-329)

Qy 6 ValGlyArgLysGlnHisPheLysValThrPheTrpGluThrAspLeuSerAsnLys 25
Db 324 GTGGGAAGAAACAGCATTTTAAAGTAACCTTTTGGGAGACTGATTGGATAATAAA 265
Qy 26 ThrLeuValSerLeuLys--LysLysLysProPheHisLeuTyrCysValIleTyrIleP 45
Db 264 ACTGTGTCCTCCCTTAGAAGAAAAAACCCTTCCACCTTACTGTGCAITATATCC 205
Qy 45 roLeuValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuI 65
Db 204 CCTAGTTCACAGTAATATATCTTATTTCTGGATATTGCTTTATACCAAGACCTT 145
Qy 65 leSer 66
Db 144 TCAGC 140

RESULT 4

BF678862

LOCUS

DEFINITION

602153173Fl NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4294170 5',

mRNA sequence.

ACCESSION

BF678862

VERSION

BF678862.1 GI:11952757

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

EST.

human.

Homo sapiens

Mammalia; Eutheria; Primates; Carnivora; Euteleostomi;

1 (Bases 1 to 615)

NIH-MGC http://mgs.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs@rmail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.

CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCML143 row: a column: 19

High quality sequence stop: 569.

Location/Qualifiers

1. 615

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:4294170"

/clone_lib="NIH_MGC_83"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: prostate; Vector: pBNR-LIB (Clontech);

Site_1: SfiI (ggcgccctgcgc); Site_2: SfiI (ggccattatggcc

); 5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGGCGCGGACATG-dt(30)BN-3'

(where B = A, C, G, or T). Average

insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

BASE COUNT 150 a 128 c 173 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 2 87e-14 Length: 615

Score: 201.00 Matches: 47

Percent Similarity: 96.00% Conservative: 1

Best Local Similarity: 94.00% Mismatches: 2

Query Match: 38.51% Indels: 2

DB: 12 Gaps: 0

US-09-830-244B-1 (1-95) x BF678862 (1-615)

Qy 17 TrpGluThrAspLeuSerAsnLysThrLeuValSerLeuLysLysLysProphe 36

Db 1 TGGGAGACTGATTG-AGTAATAATAAACTCTGCTCCCTTAAGAAAAAACCTTC 59

Qy 37 HisLeuTyrCysValIleTyrIleProLeuValProLysLeuIleLeuPheLeuAsp 56

Db 60 CACCTTTACTGTGCAITATATATCCCTTAGTCCAAAGTTAATATCTATTCTCGAT 119

Qy 57 IleAlaPheIleProLysSerLeuIleSer 66

Db 120 AT-CCTTTTATACCAAGACCTTATCAGC 148

RESULT 5

AA376683

LOCUS

DEFINITION

EST89136 Salivary gland Homo sapiens cDNA 5' end similar to

stathmin, mRNA sequence.

ACCESSION

AA376683

VERSION

AA376683.1 GI:2029001

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Dbb 211 CATACCAACCACAATACCAACAATATACCTTTTAATATCATCAGTAACCTGCAGGACA 267

```

Db 118 -----GAATTCGAAGATTCGGTTATGGGTATGGCCCTTATCAGC 156
Qy 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 157 CAGTTCCAGAACACCACTATACCCACACCACTACCAACCAACCAACCAACCAATATACCT 216
Qy 87 PheAsnIleIleSerAsnCysArgThr 95
Db 217 TTTAATATCATCAGTAAGTCACTGACGAGCA 243

RESULT 7
AA376695 341 bp mRNA linear EST 21-APR-1997
LOCUS EST89151 Salivary gland Homo sapiens cDNA 5' end similar to
DEFINITION statherin, mRNA sequence.
ACCESSION AA376695
VERSION AA376695.1 GI:2029013
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 341)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
6026280
Other ESTs: EST89150 THC77891
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse
FEATURES
Location/Qualifiers
1..341
/db_xref="ATCC (inhost):181115"
/db_xref="taxon:9606"
/clone_lib="salivary gland"
/dev_stage="adult"
/note="Organ: salivary gland; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 99 a 82 c 106 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7.49e-12 Length: 341
Score: 178.50 Matches: 37
Percent Similarity: 63.77% Conservative: 7

```

```

Best Local Similarity: 53.62% Mismatches: 8
Query Match: 34.20% Indels: 17
Db: 9 Gaps: 2
US-09-830-244B-1 (1-95) x AA376695 (1-341)
Qy 27 LeuValSerLeuLysLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 102 CTGATTTCATCGAAGCAAAAT-----TTTTCGCGTA-----131
Qy 47 ValProLysLeuIleIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
Db 132 -----GAATTCGAAGATTCGGTTATGGTATGGCCCTTATCAGC 170
Qy 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 171 CAGTTCCAGAACACCACTATACCCACCAACCAACCAACCAACCAATATACCT 230
Qy 87 PheAsnIleIleSerAsnCysArgThr 95
Db 231 TTTAATATCATCAGTAAGTCACTGACGAGCA 257

RESULT 8
AA376680 426 bp mRNA linear EST 21-APR-1997
LOCUS EST89133 Salivary gland Homo sapiens cDNA 5' end similar to
DEFINITION statherin, mRNA sequence.
ACCESSION AA376680
VERSION AA376680.1 GI:2028998
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 426)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
6026280
Other ESTs: EST89132 THC77891
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse
FEATURES
Location/Qualifiers
1..426
/db_xref="ATCC (inhost):107024"
/db_xref="taxon:9606"

```

```

/clone_lib="Salivary gland"
/dev_stage="adult"
/note="Organ: salivary gland; Vector: pBluescript SK-;
Site_1: ECORI; Site_2: XhoI"
BASE COUNT 128 a 97 c 64 g 135 t 2 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 9.74e-12 Length: 426
Score: 178.50 Matches: 37
Percent Similarity: 63.77% Conservative: 7
Best Local Similarity: 53.62% Mismatches: 8
Query Match: 34.20% Indels: 17
DB: 9 Gaps: 2

```

US-09-830-244B-1 (1-95) x AA376680 (1-426)

```

Qy 27 LeuValSerLeuLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
||||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 116 CTGATTCATCTGAAGAGAAAT-----TTTTCGGTA----- 145
Qy 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
||||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 146 -----GAATTGGAAGATTTCGGTTATGGGTATGGCCCTTATCAGC 184
Qy 67 GlnPheGlnAsnHisTyrThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
||||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 185 CAGTTCAGACACCACTATACCCACCAACCAATACCAACCAATACCAATATACCT 244
Qy 87 PheAsnIleLeuSerAsnCysArgThr 95
||||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 245 TTTAATATCATCATGTAAGTACTGAGGACA 271

```

```

RESULT 9
BG205925 321 bp mRNA linear EST 21-APR-2001
LOCUS
DEFINITION RST25360 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.
ACCESSION BG205925
VERSION BG205925.1 GI:13727612
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 321)
AUTHORS Cain, S., Leventhal, C., Thornton, M., Ramchandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Boozer, S., Mays, R., Smith,
E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher,
J., Danzig, J., and Ducar, M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
2127151
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 271.
Location/Qualifiers
1. 321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, In press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

```

FEATURES

source

```

BASE COUNT 97 a 78 c 48 g 95 t 3 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.06e-11 Length: 321
Score: 177.00 Matches: 32
Percent Similarity: 87.80% Conservative: 4
Best Local Similarity: 78.05% Mismatches: 5
Query Match: 33.91% Indels: 0
DB: 12 Gaps: 0

```

US-09-830-244B-1 (1-95) x BG205925 (1-321)

```

Qy 55 LeuAspIleAlaPheIleProLysSerLeuIleSerGlnPheGlnAsnAsnHisTyrThr 74
||||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 125 TTGGAAGATTTCGGTTATGGGTATGGCCCTTATCAGCAGTTCCAGAACACCACTATACC 184
Qy 75 HisAsnHisThrAsnHisAsnThrAsnAsnIleArgPheAsnIleIleSerAsnCysArg 94
||||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 185 CACAACCAACCAACCAACCAATACCAATATACCTTTTAAATATCATCATGTAAGTGCAGG 244
Qy 95 Thr 95
|||
Db 245 ACA 247

```

```

RESULT 10
AA376714 335 bp mRNA linear EST 21-APR-1997
LOCUS
DEFINITION EST89171 Salivary gland Homo sapiens cDNA 5' end similar to
AA376714
ACCESSION AA376714
VERSION AA376714.1 GI:2029032
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 335)
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult,
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White,
O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
L.M., Fitzhugh, W.M., Fritchman, J.B., Geoghagen, N.S., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Matmaros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Peiligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon,
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other ESTs: EST89170 THC77891
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 335

```

FEATURES

source

```

TITLE
JOURNAL
MEDLINE
COMMENT

```

BASE COUNT 109 a 75 c 43 g 107 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.47e-11 Length: 335

Score: 176.00 Matches: 31

Percent Similarity: 96.97% Conservative: 1

Best Local Similarity: 93.94% Mismatches: 1

Query Match: 33.72% Indels: 0

DB: 9 Gaps: 0

US-09-830-244B-1 (1-95) x AA376714 (1-335)

QY 63 SerLeuileSerGlnPheGlnAsnHsThrHisAsnHsThrAsnHsAsnThr 82

Db 21 GCCCTTATCAGCCAGTTCACAGAACACCTATACCCACCAACCAATACCAATACC 80

QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95

Db 81 AACAAATATACCTTTTAAATATATCATCATGTAAGTGCAGGACA 119

RESULT 11

LOCUS BG200817 470 bp mRNA linear EST 21-APR-2001

DEFINITION RST20025 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

ACCESSION BG200817

VERSION BG200817.1 GI:13722392

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.

TITLE Creation of genome-wide protein expression libraries using random activation of gene expression

JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)

MEDLINE 21227151

COMMENT Contact: Scott J. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scaine@atersys.com

High quality sequence stop: 470.

Location/Qualifiers

1..470

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Athersys RAGE Library"

/cell_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 156 a 99 c 61 g 154 t

ORIGIN

Alignment Scores:

Pred. No.: 2.19e-11 Length: 470

Score: 176.00 Matches: 36

Percent Similarity: 96.97% Conservative: 1

Best Local Similarity: 93.94% Mismatches: 1

Query Match: 33.72% Indels: 0

DB: 9 Gaps: 0

US-09-830-244B-1 (1-95) x AA376714 (1-335)

QY 63 SerLeuileSerGlnPheGlnAsnHsThrHisAsnHsThrAsnHsAsnThr 82

Db 21 GCCCTTATCAGCCAGTTCACAGAACACCTATACCCACCAACCAATACCAATACC 80

QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95

Db 81 AACAAATATACCTTTTAAATATATCATCATGTAAGTGCAGGACA 119

RESULT 11

LOCUS BG200817 470 bp mRNA linear EST 21-APR-2001

DEFINITION RST20025 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

ACCESSION BG200817

VERSION BG200817.1 GI:13722392

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.

TITLE Creation of genome-wide protein expression libraries using random activation of gene expression

JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)

MEDLINE 21227151

COMMENT Contact: Scott J. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scaine@atersys.com

High quality sequence stop: 470.

Location/Qualifiers

1..470

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Athersys RAGE Library"

/cell_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

Score: 176.00 Matches: 31

Percent Similarity: 96.97% Conservative: 1

Best Local Similarity: 93.94% Mismatches: 1

Query Match: 33.72% Indels: 0

DB: 12 Gaps: 0

US-09-830-244B-1 (1-95) x BG200817 (1-470)

QY 63 SerLeuileSerGlnPheGlnAsnHsThrHisAsnHsThrAsnHsAsnThr 82

Db 83 GCCCTTATCAGCCAGTTCACAGAACACCTATACCCACCAACCAATACCAATACC 142

QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95

Db 143 AACAAATATACCTTTTAAATATATCATCATGTAAGTGCAGGACA 181

RESULT 12

LOCUS BG188342 517 bp mRNA linear EST 21-APR-2001

DEFINITION RST7360 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

ACCESSION BG188342

VERSION BG188342.1 GI:13710029

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.

TITLE Creation of genome-wide protein expression libraries using random activation of gene expression

JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)

MEDLINE 21227151

COMMENT Contact: Scott J. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scaine@atersys.com

High quality sequence stop: 517.

Location/Qualifiers

1..517

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Athersys RAGE Library"

/cell_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 171 a 110 c 73 g 163 t

ORIGIN

Alignment Scores:

Pred. No.: 3.72e-11 Length: 517

Score: 174.50 Matches: 36

Percent Similarity: 62.32% Conservative: 7

Best Local Similarity: 52.11% Mismatches: 9

Query Match: 33.43% Indels: 17

DB: 12 Gaps: 2

US-09-830-244B-1 (1-95) x BG188342 (1-517)

QY 27 LeuValSerLeuLysLysLysLysPropPheHisLeuTyrcysValIleTyrlleProLeu 46

Db 74 CTGATTTCATCTGAAGAGAAAT-----TTTTCGCTA----- 103

Alignment Scores:

Pred. No.: 3.72e-11 Length: 517

Score: 174.50 Matches: 36

Percent Similarity: 62.32% Conservative: 7

Best Local Similarity: 52.11% Mismatches: 9

Query Match: 33.43% Indels: 17

DB: 12 Gaps: 2

US-09-830-244B-1 (1-95) x BG188342 (1-517)

```

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheLeuProLysSerLeuIleSer 66
Db 104 -----GAATTGGAAGATTCGGTTATGGGTATGGCCCTTATCAGC 142
QY 67 GluPheGlnAsnHisThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 143 CAGTTCAGAACCAACCATATACCAACCAACCAACCAACCAACCAACCAATATACCT 202
QY 87 PheAsnIleIleSerAsnCysArgThr 95
Db 203 TTTAATATCTTCAGTAAGTGCAGGACA 229

RESULT 13
LOCUS AW951480 519 bp mRNA linear EST 01-JUN-2000
DEFINITION EST363550 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
ACCESSION AW951480
VERSION AW951480.1 GI:8141155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 519)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 27
Seq primer: Reverse.
FEATURES
source
1..519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGB"
/note="Vector: pBluescriptSKm"
BASE COUNT 168 a 113 c 79 g 159 t
ORIGIN
Alignment Scores:
Pred. No.: 4.94e-11 Length: 519
Score: 173.50 Matches: 36
Percent Similarity: 63.77% Conservative: 8
Best Local Similarity: 52.17% Mismatches: 8
Query Match: 33.24% Indels: 17
DB: 10 Gaps: 2
US-09-830-244B-1 (1-95) x AW951480 (1-519)
QY 27 LeuValSerLeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46
Db 111 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 140
QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66
Db 141 -----GAATTGGAAGATTCGGTTATGGGTATGGCCCTTATCAGC 179
QY 67 GluPheGlnAsnHisThrHisAsnHisThrAsnHisAsnThrAsnAsnIleArg 86
Db 180 CAGTTCAGAACCAACCATATACCAACCAACCAACCAACCAACCAACCAATATACCT 239
QY 87 PheAsnIleIleSerAsnCysArgThr 95
Db 240 TTTAATATCTTCAGTAAGTGCAGGACA 266

```

RESULT 14
AA376682

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-830-244B-1 (1-95) x AA376682 (1-457)

QY 27 LeuValSerLeuLysLysLysLysProPheHisLeuTyrCysValIleTyrIleProLeu 46

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

QY 47 ValProLysLeuIleLeuPheLeuAspIleAlaPheIleProLysSerLeuIleSer 66

Db 102 CTGATTCATCTGAGAGAAAT-----TTTTCGCTA----- 131

```

Db 132 -----GAATTGGAAGATTCGGTTATGGGTATGGCCCTTATCAGC 170
QY 67 GlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnHisAsnHisAsnThr 82
|||||
Db 171 CAGTCCAGACACACCATATACCCACACCAACCAACCAATACCAATATACCT 230
QY 87 PheAsnIleIleSerAsnCysArgThr 95
|||
Db 231 TTTANTATCATCAGTAACCTGCAGGACA 257
RESULT 15
AA376713/c
LOCUS
DEFINITION
  AA376713 Salivary gland Homo sapiens cDNA 3' end similar to
  statherin, mRNA sequence.
ACCESSION
  AA376713
VERSION
  AA376713.1 GI:2029031
KEYWORDS
  EST.
SOURCE
  human.
  ORGANISM
    Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      1 (bases 1 to 404)
  AUTHORS
    Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
    ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
    ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
    Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
    ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
    Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
    Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
    Moreno-Palantes,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
    Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
    Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
    Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
    Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
    Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
    Kunsch,C., Hunglun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
    Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
    ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
    Venter,J.C.
  TITLE
    Initial assessment of human gene diversity and expression patterns
    based upon 83 million nucleotides of cDNA sequence
  JOURNAL
    Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE
    96026280
  COMMENT
    Other_ESTs: EST89171 THC77891
    Contact: Kerlavage, AR
    Bioinformatics
    The Institute for Genomic Research
    9712 Medical Center Drive, Rockville, MD 20850 USA
    Tel: 3018699056
    Fax: 3018699423
    Email: arkerlav@tigr.org
    For clone availability, additional sequence and expression
    information related to this EST, please check the TIGR Human Gene
    Index (http://www.tigr.org/tdb/hgi/hgi.html)
    Seq primer: M13-21.
    Location/Qualifiers
      1..404
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):181124"
        /db_xref="taxon:9606"
        /clone_lib="Salivary gland"
        /dev_stage="adult"
        /note="Organ: salivary gland; Vector: pBluescript SK-;
        Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 123 a 54 c 88 g 133 t 6 others
ORIGIN

```

```

Best Local Similarity: 87.88% Mismatches: 2
Query Match: 31.42% Indels: 0
DB: 9 Gaps: 0
US-09-830-244b-1 (1-95) x AA376713 (1-404)
QY 63 SerLeuIleSerGlnPheGlnAsnAsnHisTyrThrHisAsnHisThrAsnHisAsnThr 82
|||||
Db 376 GCCCTTATCAGCCAGGTCAGGAGCAACCACTATACCCACCAACCAATACCAATACCAATACC 317
QY 83 AsnAsnIleArgPheAsnIleIleSerAsnCysArgThr 95
|||||
Db 316 AACAAATATACCCNTTTTAAATATATCATCAGTAACCTGCAGGACA 278
Search completed: April 15, 2003, 00:02:02
Job time : 1479 secs

```

Alignment Scores:
 Pred. No.: 5.17e-10 Length: 404
 Score: 164.00 Matches: 29
 Percent Similarity: 93.94% Conservative: 2

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 20:01:06 ; Search time 3567 Seconds
(without alignments)
10859.499 Million cell updates/sec

Title: US-09-830-244B-2
Perfect score: 1331
Sequence: 1 ctatgttttagaatcaag.....gcctttccaatgtcacttg 1331

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_ov.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_br.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pln.*
- 35: em_hgt_rtd.*
- 36: em_hgt_mam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgtgo_hum.*
- 40: em_hgtgo_mus.*
- 41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	704	52.9	133984	9	AL512306	AL512306 Human DNA
C 2	396	29.8	153023	2	AC021462	AC021462 Homo sapi
C 3	361.2	27.1	153023	2	AC021462	AC021462 Homo sapi
C 4	349.2	26.2	146805	2	AL450424	AL450424 Homo sapi
C 5	253	19.0	542	9	HUMSTRNA	MI8371 Human stath
C 6	253	19.0	552	9	HUMSTRNA	MI8078 Human stath
C 7	253	19.0	1584	9	AK092678	AK092678 Homo sapi
C 8	176.2	13.2	297	11	G05711	G05711 human Srs S
C 9	145.4	10.9	1498	11	G26643	G26643 human Srs S
C 10	145.4	10.9	4723	9	HUMSTATH2	M32639 Human saliv
C 11	145.4	10.9	141568	9	AC063956	AC063956 Homo sapi
C 12	145.4	10.9	161549	2	AC024676	AC024676 Homo sapi
C 13	86.2	6.5	524	9	HUMHIS2X	M26665 Human hist
C 14	86.2	6.5	558	9	BC009791	BC009791 Homo sapi
C 15	81	6.1	480	9	HUMHIS1X	M26664 Human hist
C 16	81	6.1	566	9	BC017835	BC017835 Homo sapi
C 17	78.8	5.9	491	9	HUMBRPA	MI8372 Human hist
C 18	77.2	5.8	209410	2	AL805933	AL805933 Mus muscu
C 19	74.6	5.6	9881	6	AX281311	AX281311 Sequence
C 20	72	5.4	193445	2	AL845502	AL845502 Mus muscu
C 21	64.8	4.9	81962	2	AC128396	AC128396 Rattus no
C 22	64.8	4.9	246173	2	AC094026	AC094026 Rattus no
C 23	62.2	4.7	185969	2	AC069037	AC069037 Homo sapi
C 24	59	4.4	1764	9	HUMHIS103	L05512 Human hist
C 25	55.8	4.2	9881	6	AX281312	AX281312 Sequence
C 26	52.4	3.9	164399	3	PFMAL3P6	Z98551 Plasmodium
C 27	52.2	3.9	128635	9	AC108074	AC108074 Homo sapi
C 28	52.2	3.9	154472	2	AC109457	AC109457 Homo sapi
C 29	52.2	3.9	207558	2	AC114778	AC114778 Homo sapi
C 30	51.6	3.9	170141	9	AC092022	AC092022 Homo sapi
C 31	51.6	3.9	171044	2	AC092787	AC092787 Homo sapi
C 32	51.6	3.9	194874	2	AC080090	AC080090 Homo sapi
C 33	51.6	3.9	274349	2	AC093623	AC093623 Homo sapi
C 34	51.2	3.8	12029	3	AE001373	AE001373 Plasmodiu
C 35	51	3.8	182362	9	AL160033	AL160033 Human DNA
C 36	48.6	3.7	1141	6	AX083744	AX083744 Sequence
C 37	48.6	3.7	157051	9	AC011095	AC011095 Homo sapi
C 38	48.6	3.7	167830	2	AC027768	AC027768 Homo sapi
C 39	48.6	3.7	173765	9	AC090457	AC090457 Homo sapi
C 40	48.4	3.6	6056	6	AX345928	AX345928 Sequence
C 41	48.4	3.6	121709	9	AC093783	AC093783 Homo sapi
C 42	48.4	3.6	129989	9	AC005681	AC005681 Homo sapi
C 43	48.4	3.6	174277	9	AC093826	AC093826 Homo sapi
C 44	48	3.6	61215	2	AC090357	AC090357 Homo sapi
C 45	48	3.6	179934	9	AP000810	AP000810 Homo sapi

ALIGNMENTS

RESULT 1
AL512306/c
LOCUS Human DNA sequence from clone RP11-430C7 on chromosome 1, complete
DEFINITION 133984 bp DNA linear PRI 01-FEB-2002
ACCESSION AL512306
VERSION AL512306.16 GI:18491332
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 133984)
AUTHORS Bray-Allen, S.
TITLE Direct Submission

JOURNAL

Submitted (01-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Feb 4, 2002 this sequence version requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Swi., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chri>
Rpl1-430C7 is from the library RPCR-11.2 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-430C7. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone Rpl1-430C7 is at 133984 in this sequence. The true left end of clone Rpl1-739N20 is at 97715 in this sequence. The true right end of clone Rpl1-2317 is at 2000 in this sequence.

FEATURES

source

Location/Qualifiers

1..133984

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="Rpl1-430C7"

/clone_lib="RPCR-11.2"

123904..123933

/note="single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."

BASE COUNT 37368 a 32097 c 30202 g 34317 t

ORIGIN

Query Match 52.9%; Score 704; DB 9; Length 133984;
Best Local Similarity 99.2%; Pred. No. 1.6e-140;
Matches 718; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 1 CTATGTTTTAGAAATCAAGATGACCGGTAAAGCTGTCTCATGTACCAACAGTGAATTT 60
DB 92638 CTATGTTTTAGAAATCAAGATGACCGGTAAAGCTGTCTCATGTACCAACAGTGAATTT 92579
QY 61 ACAGTGTTCACAAATGCTGGAAATTTTGCACATGCGCATAGGGAATCTTAAGTTACTTGGC 120
DB 92578 ACAGTGTTCACAAATGCTGGAAATTTTGCACATGCGCATAGGGAATCTTAAGTTACTTGGC 92519
QY 121 TGGAAATTTATCAGCTTGTGAGTAAACAAAGTTGAGTTTACAGATGAGGGGGAATATTG 180
DB 92518 TGGAAATTTATCAGCTTGTGAGTAAACAAAGTTGAGTTTACAGATGAGGGGGAATATTG 92459
QY 181 AGGCCCTTAAGCTAAACAAATATATCAGTATCTGAGATAGTGGCTAATGTCCTCCCA 240
DB 92458 AGGCCCTTAAGCTAAACAAATATATCAGTATCTGAGATAGTGGCTAATGTCCTCCCA 92399
QY 241 GGCCTAATTTGGGAACAGTTTTTCTGATGCTTTTGAAGTACTTTCTTTTGACAGAAA 300
|||||

DB 92398 GGCCTAATTTGGGAACAGTTTTTCTGATGCTTTGAGAACTACTTCTTTTGACAGAAA 92339
QY 301 TTTTCATTTCTGCTTCCCATTTGCTATATTTCTCCCTTTATAGAGCAATTTCTTTTCC 360
DB 92338 TTTTCATTTCTGCTTCCCATTTGCTATATTTCTCCCTTTATAGAGCAATTTCTTTTCC 92279
QY 361 TTTTCGCGAAATGTCCTCATTTAGCATTTTTCAGATCTTTTGGATGTCGCACTAATGCAATAT 420
DB 92278 TTTTCGCGAAATGTCCTCATTTAGCATTTTTCAGATCTTTTGGATGTCGCACTAATGCAATAT 92219
QY 421 TGGTAATCCCTTTATTTGGTGAATACAGCATAGTTTAAATAAATCTTTACAGTAAATCTTACA 480
DB 92218 TGGTAATCCCTTTATTTGGTGAATACAGCATAGTTTAAATAAATCTTTACAGTAAATCTTACA 92159
QY 481 CTGGAATTTGCTGCACCTCTACCAATAGCCCTTTTGAATGACTGAAAGTGTTAACAGAGAA 540
DB 92158 CTGGAATTTGCTGCACCTCTACCAATAGCCCTTTTGAATGACTGAAAGTGTTAACAGAGAA 92099
QY 541 AGAGCATGCTCTGCAGAAAGAGATAGCTAATATTTTGGTACTTTATCTGAAATCCCAAG 600
DB 92098 AGAGCATGCTCTGCAGAAAGAGATAGCTAATATTTTGGTACTTTATCTGAAATCCCAAG 92039
QY 601 ATGCTGCTTCCCTGCGAGGTTGTTTCTCTTACGATCCCTCAATGTAATCCCTCTCGGA 660
DB 92038 ATGCTGCTTCCCTGCGAGGTTGTTTCTCTTACGATCCCTCAATGTAATCCCTCTCGGA 91979
QY 661 GCACAGACAGTATAGTAGAATCTCCATTTCTTTGTTTCTTTTAAAGACAGAGACTCT 720
DB 91978 GCACAGACAGTATAGTAGAATCTCCATTTCTTTGTTTCTTTTAAAGACAGAGACTCT 720
QY 721 GTCT 724
DB 91919 CTCT 91916
|||||

RESULT 2

AC021462/c

LOCUS

DEFINITION

Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered pieces.

AC021462

AC021462.3 GI:7387343

HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 153023)

Birten,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-276C1

Unpublished

2 (bases 1 to 153023)

Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lied,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 3, 2000 this sequence version replaced gi:7230200.

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5201
 Center clone name: 276_C.1
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 141815 bases at least Q40
 Consensus quality: 147368 bases at least Q30
 Consensus quality: 149481 bases at least Q20
 Insert size: 157000; agarose-fp
 Insert size: 151023; sum-of-contigs
 Quality coverage: 4.3 in Q20 bases; agarose-fp
 Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1016 1115: contig of 1015 bp in length
 1116 2276: contig of 1161 bp in length
 2277 2376: gap of 100 bp
 2377 4009: contig of 1633 bp in length
 4010 4109: gap of 100 bp
 4110 5737: contig of 1628 bp in length
 5738 5837: gap of 100 bp
 5838 6939: contig of 1102 bp in length
 6940 7039: gap of 100 bp
 7040 9504: contig of 2465 bp in length
 9505 9604: gap of 100 bp
 9605 11734: contig of 2130 bp in length
 11735 11834: gap of 100 bp
 11835 14111: contig of 2277 bp in length
 14112 14211: gap of 100 bp
 14212 16973: contig of 2762 bp in length
 16974 17073: gap of 100 bp
 17074 19588: contig of 2515 bp in length
 19589 19688: gap of 100 bp
 19689 22275: contig of 2587 bp in length
 22276 22375: gap of 100 bp
 22376 25823: contig of 3448 bp in length
 25824 25923: gap of 100 bp
 25924 31307: contig of 5384 bp in length
 31308 31407: gap of 100 bp
 31408 34863: contig of 3456 bp in length
 34864 34963: gap of 100 bp
 34964 41383: contig of 6419 bp in length
 41383 41482: gap of 100 bp
 41483 49024: contig of 7542 bp in length
 49025 49124: gap of 100 bp
 49125 58672: contig of 9548 bp in length
 58673 58772: gap of 100 bp
 58773 69623: contig of 10850 bp in length
 69623 69722: gap of 100 bp
 69723 88191: contig of 18469 bp in length
 88192 88291: gap of 100 bp
 88292 107084: contig of 18793 bp in length
 107085 107184: gap of 100 bp
 107185 153023: contig of 45839 bp in length.

FEATURES

source
 1. .153023
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-276C1"
 /clone_lib="RPC1-11 Human Male BAC"
 1. .1015
 /note="assembly_fragment"
 1116. 2276
 /note="assembly_fragment"
 2377. .4009
 /note="assembly_fragment"
 4110. .5737
 /note="assembly_fragment"
 5838. 6939
 /note="assembly_fragment"
 7040. .9504
 /note="assembly_fragment"
 9605. .11734
 /note="assembly_fragment"
 11835. 14111
 /note="assembly_fragment"
 14212. .16973
 /note="assembly_fragment"
 17074. .19588
 /note="assembly_fragment"
 19689. 22275
 /note="assembly_fragment"
 22376. .25823
 /note="assembly_fragment"
 25924. .31307
 /note="assembly_fragment"
 31408. 34863
 /note="assembly_fragment"
 34964. .41382
 /note="assembly_fragment"
 41483. .49024
 /note="assembly_fragment"
 49125. 58672
 /note="assembly_fragment"
 58773. .69622
 /note="assembly_fragment"
 69723. .88191
 /note="assembly_fragment"
 88292. .107084
 /note="assembly_fragment"
 107185. .153023
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left
 BASE COUNT 41494 a 35192 c 34573 g 39756 t 2008 others
 ORIGIN
 Query Match 29.8%; Score 396; DB 2; Length 153023;
 Best Local Similarity 100.0%; Pred. No. 8.1e-75;
 Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTATGTTTTTGTAGATCAAGATGAACCGGTAAGCTGCTCATGTACCAACAGTGAATTT 60
 Db 4505 CTATGTTTTTGTAGATCAAGATGAACCGGTAAGCTGCTCATGTACCAACAGTGAATTT 4446
 QY 61 ACAGTGTTTTACAAATGCTGGAATTTTGCACCTGCATAGGGAATTTTAAAGGTTACTTGGC 120
 Db 4445 ACAGTGTTTTACAAATGCTGGAATTTTGCACCTGCATAGGGAATTTTAAAGGTTACTTGGC 4386
 QY 121 TGAATTTATCAGACTTGTGAGTAAACAAGTTTACAGATGAGGGGAATATTG 180
 Db 4385 TGAATTTATCAGACTTGTGAGTAAACAAGTTTACAGATGAGGGGAATATTG 4326
 QY 181 AGCCCCCTTAGGCTAAACAATAATCAAGTATCTGAGATAGTGGCTAATGTGGCTCCCA 240
 Db 4325 AGCCCCCTTAGGCTAAACAATAATCAAGTATCTGAGATAGTGGCTAATGTGGCTCCCA 4266
 QY 241 GGCTAATTTGGGAACAGTTTTTCTCCTGATTGCTTTGAGAGAGTACTTCTTTGACAGAA 300

```

Db 4265 GGCCTAATTTGGGAACAGATTTTCTGATGCTTTGAGAAGTACATTTCTTTTGACAGAA 4206
Qy 301 TTTTCATTTCTGCTGCCATTTGCTATATCTCCCTTTATAGAGCATTGGATTTCTTTCC 360
Db 4205 TTTTCATTTCTGCTGCCATTTGCTATATCTCCCTTTATAGAGCATTGGATTTCTTTCC 4146
Qy 361 TTTTGTGGGAATGTCCTCATTTAGCATTTTCAGATCT 396
Db 4145 TTTTGTGGGAATGTCCTCATTTAGCATTTTCAGATCT 4110

RESULT 3
AC021462
LOCUS Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
AC021462
AC021462.3 GI:7387343
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 153023)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-276C1
Unpublished
2 (bases 1 to 153023)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepli,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArillano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
McDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Melidrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7230200.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5201
Center clone name: 276_C_1
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141815 bases at least Q40
Consensus quality: 147368 bases at least Q30
Consensus quality: 149481 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 151023; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 1015: contig of 1015 bp in length
* 1016 1115: gap of 100 bp
* 1116 2276: contig of 1161 bp in length
* 2277 2376: gap of 100 bp
* 2377 4009: contig of 1633 bp in length
* 4010 4109: gap of 100 bp
* 4110 5737: contig of 1628 bp in length
* 5738 5837: gap of 100 bp
* 5838 6939: contig of 1102 bp in length
* 6940 7039: gap of 100 bp
* 7040 9504: contig of 2465 bp in length
* 9505 9604: gap of 100 bp
* 9605 11734: contig of 2130 bp in length
* 11735 11834: gap of 100 bp
* 11835 14111: contig of 2277 bp in length
* 14112 14211: gap of 100 bp
* 14212 16973: contig of 2762 bp in length
* 16974 17073: gap of 100 bp
* 17074 19588: contig of 2515 bp in length
* 19589 19688: gap of 100 bp
* 19689 22275: contig of 2587 bp in length
* 22276 22375: gap of 100 bp
* 22376 25823: contig of 3448 bp in length
* 25824 25923: gap of 100 bp
* 25924 31307: contig of 5384 bp in length
* 31308 31407: gap of 100 bp
* 31408 34863: contig of 3456 bp in length
* 34864 34963: gap of 100 bp
* 34964 41382: contig of 6419 bp in length
* 41383 41482: gap of 100 bp
* 41483 49024: contig of 7542 bp in length
* 49025 49124: gap of 100 bp
* 49125 58672: contig of 9548 bp in length
* 58673 58772: gap of 100 bp
* 58773 69622: contig of 10850 bp in length
* 69623 69722: gap of 100 bp
* 69723 88191: contig of 18469 bp in length
* 88192 88291: gap of 100 bp
* 88292 107084: contig of 18793 bp in length
* 107085 107184: gap of 100 bp
* 107185 153023: contig of 45839 bp in length.

```

FEATURES

source

```

1. .153023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-276C1"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature
1. .1015
/note="assembly_fragment"

misc_feature
1116. .2276
/note="assembly_fragment"

misc_feature
2377. .4009
/note="assembly_fragment"

misc_feature
4110. .5737
/note="assembly_fragment"

misc_feature
5838. .6939
/note="assembly_fragment"

misc_feature
7040. .9504
/note="assembly_fragment"

misc_feature
9605. .11734
/note="assembly_fragment"

misc_feature
11835. .14111
/note="assembly_fragment"

misc_feature
14212. .16973
/note="assembly_fragment"

```


* 107258 110300: contig of 3043 bp in length
 * 110301 110400: gap of 100 bp
 * 110401 113955: contig of 3555 bp in length
 * 113956 114055: gap of 100 bp
 * 114056 118801: contig of 4746 bp in length
 * 118802 118901: gap of 100 bp
 * 118902 129831: contig of 10930 bp in length
 * 129832 129931: gap of 100 bp
 * 129932 137545: contig of 7614 bp in length
 * 137546 137645: gap of 100 bp
 * 137646 146805: contig of 9160 bp in length.

FEATURES

source
 1. 146805
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-563I16"
 /clone.lib="RPC1-11.2"
 1. 5105
 /note="assembly_fragment:00368
 clone_end:T7
 vector_side:left"
 5206.10570
 /note="assembly_fragment:00594
 fragment_chain:1"
 10671..17626
 /note="assembly_fragment:00786
 fragment_chain:1"
 17727..20478
 /note="assembly_fragment:00844
 fragment_chain:1"
 20579..23035
 /note="assembly_fragment:00011
 fragment_chain:1"
 23136..25308
 /note="assembly_fragment:01029
 fragment_chain:1"
 25409..30469
 /note="assembly_fragment:00148
 fragment_chain:2"
 30570..32688
 /note="assembly_fragment:00961
 fragment_chain:2"
 32789..36187
 /note="assembly_fragment:00310
 fragment_chain:3"
 36288..48988
 /note="assembly_fragment:01150
 fragment_chain:3"
 49089..67333
 /note="assembly_fragment:00444
 fragment_chain:4"
 67434..82317
 /note="assembly_fragment:00140
 fragment_chain:4"
 82419..84614
 /note="assembly_fragment:00657
 fragment_chain:5"
 84715..90326
 /note="assembly_fragment:00757
 fragment_chain:5"
 90427..92847
 /note="assembly_fragment:00677
 fragment_chain:6"
 92748..97690
 /note="assembly_fragment:00248
 fragment_chain:6"
 97791..100799
 /note="assembly_fragment:00091"
 100900..104202
 /note="assembly_fragment:00538"
 104303..107157
 /note="assembly_fragment:00675"

misc_feature 107258..110300
 /note="assembly_fragment:00695"
 110401..113955
 /note="assembly_fragment:01316"
 114056..118801
 /note="assembly_fragment:01292
 fragment_chain:7"
 118902..129831
 /note="assembly_fragment:00551
 fragment_chain:7"
 129932..137545
 /note="assembly_fragment:01410
 fragment_chain:7"
 137646..146805
 /note="assembly_fragment:00763
 fragment_chain:7"
 clone_end:SP6
 vector_side:right"
 BASE COUNT 37508 a 33240 c 34076 g 39557 t 2424 others
 ORIGIN

Query Match 26.2%; Score 349.2; DB 2; Length 146805;
 Best Local Similarity 96.3%; Pred. No. 7.9e-65;
 Matches 368; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
 QY 709 GACAGAGACTCTGCTCAAAAAAGGACATTTATCATATATACATCTTATTAGAGCCCC 768
 Db 84740 GACAGAGACTCTGCTCAAAAAAGGACATTTATCATATATACATCTTATTAGAGCCCC 84799
 QY 769 TAATTTCTATCTGAAGGCACCTGTTTTTTTTTTTAAACAGTTAAGTACTGATCTCAACAG 828
 Db 84800 TAATTTCTATCTGAAGGCACCTGTTTTTTTTTTTAAACAGTTAAGTACTGATCTCAACAG 84859
 QY 829 ACAATATTTCTGATCAGATAGTCCCTCTCAACAGTAGCAATGTGGTTCATAAAGTG 888
 Db 84860 ACAATATTTCTGATCAGATAGTCCCTCTCAACAGTAGCAATGTGGTTCATAAAGTG 84919
 QY 889 GGAAGAAACAGCATTTTAAAGTAACTTTTGGGAGACTGATTGAGTAAATAAACT 948
 Db 84920 GGAAGAAACAGCATTTTAAAGTAACTTTTGGGAGACTGATTGAGTAAATAAACT 84978
 QY 949 CTGGTCTCCCTTAAGAAAAAACCCCTTCCACCTTTACTGTCTATTTATATCCCTTA 1008
 Db 84979 CTGGTCTCCCTTAAGAAAAAACCCCTTCCACCTTTACTGTCTATTTATATCCCTTA 85038
 QY 1009 GTTCCAAAGTTAATTATCTTATTTCTGGATATGCTTTTATACCAAGAGCCCTTATCAGC 1068
 Db 85039 GTTCCAAAGTTAATTATCTTATTTCTGGATATGCTTTTATACCAAGAGCCCTTATCAGC 85098
 QY 1069 CAGTTCAGACAACACACTATA 1090
 Db 85099 CCTTGTAACATACAGTATCTTTA 85120

RESULT 5
 HUMSTTRNA
 LOCUS Human statherin mRNA, complete cds.
 DEFINITION Human statherin mRNA, complete cds.
 ACCESSION M18371
 VERSION M18371.1 GI:338610
 KEYWORDS statherin.
 SOURCE Human female submandibular gland, cDNA to mRNA, clone pBRHSMF988.2.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 542)
 AUTHORS Dickinson,D.P., Ridall,A.L. and Levine,M.J.
 TITLE Human submandibular gland statherin and basic histidine-rich
 peptide are encoded by highly abundant mRNA's derived from a common
 ancestral sequence
 JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)
 MEDLINE 88106506
 PUBMED 3426601

Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 1584)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: HRI and

RAB; annotation: HRI and RAB.

Location/Qualifiers

1. 1584

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="SALGL100107"

/tissue_type="salivary gland"

/clone_lib="SALGL1"

/notes="cloning vector: pME18SFL3"

89. .997

/notes="unnamed protein product"

/codon_start=1

/protein_id="BAC03943.1"

/db_xref="GI:21751328"

/translation="MGDDSEMLKLPVDQCEHLKWLKARLSGYEEALKIFQIKNDKSP

EWKFLGLTKFFVDSNVAQLKGLAIVYENAHVAGTKTGEVTVGVSVKFNOPK

AKAEGLIEICLMYIEIEGEAVQEEELKGLNPKIIVACITLTKALSRFGKII

LKPIIKVLPLKESKAVRDEAKLIAVEIYRWIDALRPPLONINSVOLKELEEW

VKLPTSAIPRTFLRSQOELEAKLEQQSQSGDGEVDDGDDVPOIDAYELLEAVEI

LSKLPTFTFKLQKNGKREKRWSL"

BASE COUNT 526 a 238 c 331 g 429 t

ORIGIN

Query Match 19.0%; Score 253; DB 9; Length 1584;
Best Local Similarity 88.4%; Pred. No. 6.6e-44;
Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1022 TTATCTTATTCGGATATTCGTTTATACCAAGAGCCCTTATCAGCCAGTTCAGAAC 1081

Db 1174 TTTGCGTAGAATTGGAAGATTCGGTTATGGGTATGGCCCTTATCAGCCAGTTCAGAAC 1233

QY 1082 ACCACTATACGCAACCAATACCAACCAATACCAATACCAATATACGTTTAAATATCATCA 1141

Db 1234 ACCACTATACCAACCAATACCAACCAATACCAATACCAATATACGTTTAAATATCATCA 1293

QY 1142 GTAACCTCAGGACATGATTATTGAGGCTTGATTGGCAATACGACTTCTACATCCATATT 1201

Db 1294 GTRAACCTCAGGACATGATTATTGAGGCTTGATTGGCAATACGACTTCTACATCCATATT 1353

QY 1202 CTATCTTTTATACCATATACCATATACCATATTTTGTGAGATCATCTPAAGAGCAATG 1261

Db 1354 CTATCTTTTATACCATATACCATATACCATATTTTGTGAGATCATCTPAAGAGCAATG 1413

QY 1262 CGAATGFAAAACCTTAATTTTACTGGATCTCTTGGTCCAGATACCTGCTTTTCCA 1321

Db 1414 CAATGNAACACTATATTTTACTGTATCTCTTGTGTTTTCAGGATACCTGCTTTTCCA 1473

QY 1322 ATGTCACCTTG 1331

Db 1474 TTGTCACCTTG 1483

RESULT 8

G05711

LOCUS

G05711 297 bp DNA linear STS 19-OCT-1995

DEFINITION

human STS WI-7844, sequence tagged site.

ACCESSION G05711

VERSION G05711.1 GI:858956

KEYWORDS STS; STS sequence; primer; sequence tagged site.

SOURCE Homo sapiens STSs derived from sequences in dbEST and the Unigene collection.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 297)

AUTHORS Hudson, T.

TITLE Whitehead Institute/MIT Center for Genome Research; Physically

JOURNAL Mapped ESTs

COMMENT Unpublished (1995)

Contact: Thomas Hudson

Whitehead Institute/MIT Center for Genome Research

Whitehead Institute for Biomedical Research

9 Cambridge Center, Cambridge MA 02142 USA

Tel: 617 252 1900

Fax: 617 252 1902

Email: thudson@genome.wi.mit.edu

Primer A: ACTGACGACATGATATTGAGG

Primer B: AAAATGTGCTTGAATTTATTGC

STS size: 286

PCR profile:

Presoak:

Denaturation:

Annealing: 56 degrees C

Polymerization:

PCR Cycles: 35

Thermal Cycler:

Protocol:

Template: 10 ng

Primer: each 5 pm

dNTPs: each 4 mM

Taq polymerase: 0.025 units/ul

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 9.3

Prepared with primer pairs derived from M18371 -- Unigene.

Location/Qualifiers

1. .297

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="748_F.3; 793_E.2; 794_A.9; 886_B.3; 960_D.2;

972_E.4; 750_D.11; 855_E.12; 882_G.11; 931_G.10; 395.8 cr

from top of Chr4 linkage group"

12. .297

primer_bind 12. .34

BASE COUNT 99 a 56 c 37 g 105 t

ORIGIN

Query Match 13.2%; Score 176.2; DB 11; Length 297;

Best Local Similarity 92.9%; Pred. No. 2.2e-27;

Matches 184; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1134 TATCATCACTAATCGAGCAGCATGATTATTGAGGCTTGATGGCAATACGACTTCTACA 1193

Db 1 TATCATCACTAATCGAGCAGCATGATTATTGAGGCTTGATGGCAATACGACTTCTACA 60

QY 1194 TCCATATTCCTATCTTTCATACCATATCACACTACTACCATCTTTTGTNAGATCATCTAA 1253

Db 61 TCCATATTCCTATCTTTCATACCATATCACACTACTACCATCTTTTGTNAGATCATCTAA 120

QY 1254 GAGCAATGCGAATGTAAACCCCTATTAATTTACTGGATCTCTTTGGTCCAGATCTGC 1313

Db 121 GAGCAATGCAATGAAAACACTAATATTTACTGTATACTCTTTGTTTCAGGATCTGC 180
 Oy 1314 CTTTCCAATGTCACTTG 1331
 Db 181 CTTTCCAATGTCACTTG 198

RESULT 9
 G26643
 LOCUS human STS STS_M32639, sequence tagged site.
 DEFINITION
 accession G26643
 version G26643.1 GI:1348875
 keywords STS; STS sequence; primer; sequence tagged site.
 source Homo sapiens STSs derived from sequences in dbEST and the Unigene collection.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS 1 (bases 1 to 1498)
 TITLE Hudson, T.
 JOURNAL Whitehead Institute/MIT Center for Genome Research; Physically
 COMMENT Mapped STSs
 Unpublished (1995)

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: GGACATGATTATTGAGGTAAGATGG
 Primer B: ACTGCCTTCTTAACCATAGCAGG
 STS size: 210
 PCR profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

FEATURES
 source Derived from dbEST (genbank accession M32639).
 Location/Qualifiers
 1..1498
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="473.8 CR from top of Chr4 linkage group"
 21..230
 21..45
 primer_bind complement(209..230)
 primer_bind 229 C 244 G 514 T
 BASE COUNT 511 a 229 c 244 g 514 t
 ORIGIN.

Query Match 10.9%; Score 145.4; DB 11; Length 1498;
 Best Local Similarity 90.1%; Pred. No. 6.1e-21;
 Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

exon 2728..>2814
/note="statherin precursor"
intron number=5
polya_signal /note="5'UTR Intron E"
BASE COUNT 1570 a 797 c 722 g 1634 t
ORIGIN About 1.8 kb after segment 1.

Query Match 10.9% Score 145.4; DB 9; Length 4723;
Best Local Similarity 90.1%; Pred. No. 4.8e-21;
Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1160 TATTGAGCTTGTATGGCAATACGACTTCTACATCCATATCTCATCTTTTCATACCATTA 1219
DB 4045 TGTGAGGCTTGTATGGCAATACGACTTCTACATCCATATCTCATCTTTTCATACCATTA 4104
QY 1220 TCACACTACTACACTTTTGTGAGCAATCTTAAGAGCAATGCGAATGTAACCACTATA 1279
DB 4105 TCACACTACTACACTTTTGTGAGCAATCTTAAGAGCAATGCGAATGTAACCACTATA 4164
QY 1280 ATTACTGCATCTCTTTGGTTCAGATCTTTCCTTTTCCAAATGTCACCTG 1331
DB 4165 ATTACTGCATCTCTTTGGTTCAGATCTTTCCTTTTCCAAATGTCACCTG 4216

RESULT 11
AC063956 141568 bp DNA linear PRI 25-AUG-2000
LOCUS Homo sapiens 4 BAC RP11-529K3 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
AC063956
AC063956.7 GI:9910030
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 141568)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blum,M., Blankensburg,K., Bonnin,D., Bouck,J.,
Bowling,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M.,
Holloway,C., Hollins,B., Homsai,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C.,
Liu,J., Liu,W., Loulseyed,H., Lozano,R.J., Lu,X., Lucier,A.,
Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,
Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T.,
Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Zuchowicz,R.,
Naylor,S.D., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 141568)
Worley,K.C.
Direct Submission
Submitted (22-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 141568)
Worley,K.C.
Direct Submission
Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 25, 2000 this sequence version replaced gi:9795448.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality/info/genbank.annotation.html>.

QUALSTAT-REPORT-----
----- Summary Statistics -----
Contig length: 141568
Phrap values in estimate: 140895
Average error rate (ECM-Phrap estimate): 1.4137e-05
Fraction of Phrap values less than 40 : 0.00715426
Number of consensus changing edits: 10
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
23165 cctgaattct(n)tcttgtaa cctgaattct(t)tcttgtaa


```

||||| ATTACGTATACCTTTGTTTCAGGATACCTTGCCTTTTCAATGTCACTTG 126747
||||| HUMHIS2X 524 bp mRNA linear PRI 31-DEC-1994
LOCUS Human histatin 2 (HIS2) mRNA, complete cds.
DEFINITION M26665
ACCESSION M26665.1 GI:292145
KEYWORDS histatin 2.
SOURCE Homo sapiens parotid gland cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Sabatini, L.M. and Azen, E.A.
AUTHORS 1 (bases 1 to 524)
TITLE Histatins, a family of salivary histidine-rich proteins, are
encoded by at least two loci (HIS1 and HIS2)
JOURNAL Biochem. Biophys. Res. Commun. 160 (2), 495-502 (1989)
MEDLINE 89246491
PUBMED 2719677
FEATURES
    source
        Location/Qualifiers
            1..524
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /tissue_type="parotid gland"
            1..524
                /gene="HIS2"
            72..227
                /gene="HIS2"
            /codon_start=1
            /product="histatin 2"
            /protein_id="AAA58646.1"
            /db_xref="GI:292146"
            /translation="MKFFVFALILALMLSMGTADSHAKRHHGKFKFHKHHSHRGYR
                SNLYDN"
            506..511
                /polya_signal
BASE COUNT 157 a 100 c 87 g 180 t
ORIGIN
Query Match 6.5%; Score 86.2; DB 9; Length 524;
Best Local Similarity 62.5%; Pred. No. 3.2e-08;
Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;
QY 1093 CACACCAATACCAACCAATACCAATATACGTTTAAATATCATCATCACTGTCAGG 1152
|||||
DB 187 CACATCGAGGCTATAGATCAAAATATCTGTATGACAATGTATCTTCAGTAATCATGGG 246
|||||
QY 1153 ACATGATTATGAGGCTTGATTGGCAAAATACGACTTCTACATCCATATTTCTCATCTTCA 1212
|||||
DB 247 GCATGATTATGAGGCTTGACTGGCAAAATTCGCTTTGGACTCGTGTATCTCATTTGCA 306
|||||
QY 1213 TACCATATCACACTACTACCACTTTTGTNAGATCATCTAAGAGCAATGC-GAATGTAAA 1271
|||||
DB 307 TACCGCATCACACTACCACTGCTTTTGAAGAATTTATCATAGGCAATGCAGAAATAAG 366
|||||
QY 1272 ACCCTATATTTACGGTACTCTTTGGTTCCAGATATCTGCTTTTCCATGTCACCTTG 1331
|||||
DB 367 AATACCATGATTATAGTGAATCTGTGTTTCAGGATATCTCCCTTCCCTAATATCATTTG 426
|||||
RESULT 14
BC009791
LOCUS Homo sapiens, histatin 3, clone MGC:13578 IMAGE:4293405, mRNA,
DEFINITION complete cds.
ACCESSION BC009791
VERSION BC009791.1 GI:14602560
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

```

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

CDS

BASE COUNT 187 a 102 c 91 g 178 t

ORIGIN

Query Match 6.5%; Score 86.2; DB 9; Length 558;
 Best Local Similarity 62.5%; Pred. No. 3.2e-08;
 Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1093 CACACCAATACCAACCAATACCAATATACGTTTAAATATCATCATCACTGTCAGG 1152
 |||||
 DB 191 CACATCGAGGCTATAGATCAAAATATCTGTATGACAATGTATCTTCAGTAATCATGGG 250
 |||||
 QY 1153 ACATGATTATGAGGCTTGATTGGCAAAATACGACTTCTACATCCATATTTCTCATCTTCA 1212
 |||||
 DB 251 GCATGATTATGAGGCTTGACTGGCAAAATTCGCTTTGGACTCGTGTATCTCATTTGCA 310
 |||||
 QY 1213 TACCATATCACACTACTACCACTTTTGTNAGATCATCTAAGAGCAATGC-GAATGTAAA 1271
 |||||
 DB 311 TACCGCATCACACTACCACTGCTTTTGAAGAATTTATCATAGGCAATGCAGAAATAAG 370
 |||||
 QY 1272 ACCCTATATTTACGGTACTCTTTGGTTCCAGATATCTGCTTTTCCATGTCACCTTG 1331
 |||||
 DB 371 AATACCATGATTATAGTGAATCTGTGTTTCAGGATATCTCCCTTCCCTAATATCATTTG 430
 |||||

RESULT 15
 HUMHISX
 LOCUS
 DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.
 ACCESSION M26664
 VERSION M26664.1 GI:292143

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Strausberg, R.
 Direct Submission
 Submitted (02-JUL-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-remail.nih.gov
 Tissue procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 Contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 19 Row: O Column: 1
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA GI: 4557652.

Location/Qualifiers
 1..558
 /organism="Homo sapiens"
 /db_xref="LocusID:3347"
 /db_xref="taxon:9606"
 /clone="MGC:13578 IMAGE:4293405"
 /tissue_type="Skeletal Muscle"
 /clone_lib="NIH_MGC_81"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 76..231
 /codon_start=1
 /product="histatin 3"
 /protein_id="AAH09791.1"
 /db_xref="GI:14602561"
 /translation="MKFFVFALILALMLSMGTADSHAKRHHGKFKFHKHHSHRGYR
 SNLYDN"

BASE COUNT 187 a 102 c 91 g 178 t

ORIGIN

Query Match 6.5%; Score 86.2; DB 9; Length 558;
 Best Local Similarity 62.5%; Pred. No. 3.2e-08;
 Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1093 CACACCAATACCAACCAATACCAATATACGTTTAAATATCATCATCACTGTCAGG 1152
 |||||
 DB 191 CACATCGAGGCTATAGATCAAAATATCTGTATGACAATGTATCTTCAGTAATCATGGG 250
 |||||
 QY 1153 ACATGATTATGAGGCTTGATTGGCAAAATACGACTTCTACATCCATATTTCTCATCTTCA 1212
 |||||
 DB 251 GCATGATTATGAGGCTTGACTGGCAAAATTCGCTTTGGACTCGTGTATCTCATTTGCA 310
 |||||
 QY 1213 TACCATATCACACTACTACCACTTTTGTNAGATCATCTAAGAGCAATGC-GAATGTAAA 1271
 |||||
 DB 311 TACCGCATCACACTACCACTGCTTTTGAAGAATTTATCATAGGCAATGCAGAAATAAG 370
 |||||
 QY 1272 ACCCTATATTTACGGTACTCTTTGGTTCCAGATATCTGCTTTTCCATGTCACCTTG 1331
 |||||
 DB 371 AATACCATGATTATAGTGAATCTGTGTTTCAGGATATCTCCCTTCCCTAATATCATTTG 430
 |||||

RESULT 15
 HUMHISX
 LOCUS
 DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.
 ACCESSION M26664
 VERSION M26664.1 GI:292143

Search completed: April 14, 2003, 22:34:32
Job time : 4275 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 19:59:10 ; Search time 330 Seconds

(without alignments)
9083.060 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 catgttttagaatcaag.....gcctttccaatgtcacttg 1331

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq 101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1330	99.9	1331	21	Human lysine-rich
2	253	19.0	552	21	Human scatherin DN
3	209.6	15.7	2121	21	Human ORFX ORF2220
4	187.4	14.1	221	21	Human secreted exp
5	86.6	6.5	390	22	Human polynucleoti
6	86.2	6.5	857	22	Human EST-derived
7	78.8	5.9	21	AAA48965	Human basic histid
8	78.6	5.9	126	16	Human gene signatu
9	76.4	5.7	90	16	Human gene signatu

10	74.6	5.6	9881	24	ABL54353
11	60.2	4.5	516	20	AAV89525
12	56.4	4.2	60	24	ABN43273
13	55.8	4.2	9881	24	ABL54354
14	48.4	3.6	6056	24	ABL33026
15	47	3.5	61020	22	AA546787
16	46.8	3.5	10279	24	ABL92276
17	46.8	3.5	10279	24	ABL33590
18	46.8	3.5	10279	24	AA22327
19	46.8	3.5	513445	22	AA161373
20	46	3.5	6131	24	ABL32891
21	44.8	3.4	545	22	ABL70126
22	44.6	3.4	556	22	AA191773
23	44	3.3	16766	24	ABL34157
24	43.8	3.3	5461	24	ABQ66977
25	43.8	3.3	5461	24	ABL32998
26	43.8	3.3	5647	24	ABL70355
27	43.8	3.3	5647	24	ABL33566
28	43.8	3.3	5647	24	AA563320
29	43.4	3.3	8577	24	ABL33787
30	43.4	3.3	19659	24	ABL32766
31	42.8	3.2	375	23	ABV44911
32	42.8	3.2	8946	24	ABL32910
33	42.6	3.2	3680	22	AAH14565
34	42.6	3.2	7823	22	AAH72614
35	42.4	3.2	18183	22	AA546279
36	42.4	3.2	18183	24	ABL70111
37	42.4	3.2	18183	24	ABK31158
38	42.2	3.2	1160	19	AAV44863
39	42.2	3.2	1160	22	AAF98460
40	42.2	3.2	6055	22	AA536246
41	42.2	3.2	7809	22	AA536244
42	42.2	3.2	7809	22	AA536245
43	42	3.2	707	22	AA124056
44	42	3.2	15373	24	ABL32466
45	41.8	3.1	7171	24	ABN80216

ALIGNMENTS

RESULT 1

AAA48963

ID AAA48963 standard; cDNA; 1331 BP.

XX
AC AAA48963;

XX
DT 06-OCT-2000 (first entry)

XX
DE Human lysine-rich statherin cDNA from Incyte clone 2820214.

XX
KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
KW precipitation inhibitor; autoimmunity; inflammatory disorder; AIDS;
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
KW cancer; leukemia; adenocarcinoma; melanoma; ss.

OS Homo sapiens.

XX
FH Key Location/Qualifiers
CDS 871..1158

FT /*tag= a
FT /product= Lysine-rich_statherin

XX
PN WO200024779-A1.

XX
PD 04-MAY-2000.

XX
PF 22-OCT-1999; 99WO-US24046.

XX
PR 23-OCT-1998; 98US-0155209.

XX
PA (INCY-) INCYTE PHARM INC.

XX

Chemically treated
EST clone Cp289.
Human spliced tran
Chemically treated
Human immune syste
Tumour suppressor
Chemically treated
Human immune syste
Chemically treated
Soybean 318013 reg
Human immune syste
Human cervical can
Human polynucleoti
Human immune syste
Human anglogenesis
Human immune syste
Chemically treated
Human immune syste
Human gene regulat
Human immune syste
Human prostate exp
Human immune syste
Human cDNA sequenc
Tumour suppressor
Chemically treated
Signal transductio
Clone CK201-1 codi
Human cDNA clone C
Human cardiovascular
Human cardiovascular
Human breast cance
Human immune syste
Human chemically m

PI Tang YT, Corley NC, Guegler KJ, Patterson C;
 XX WPI: 2000-350699/30.
 DR P-PSDB; AAY94526.
 XX
 PT Purified polypeptide used for treating or preventing a disorder
 PT characterized by expression or activity of lysine-rich statherin
 PT proteins -
 XX
 PS Claim 3; Page 70; 75pp; English.
 XX
 CC The present sequence is human lysine-rich statherin protein (LRSP)
 CC cDNA from Incyte clone 2820214. This sequence was identified
 CC through analysis of a cDNA library of breast tumour tissue
 CC (BRSTN014). The LRSP sequence was found to have homology with
 CC human statherin (AAY94527) and human basic histidine-rich protein
 CC (AAY94528). Human statherin is a phosphoprotein that acts as an
 CC inhibitor of precipitation of calcium phosphate salts in the oral
 CC cavity. The LRSP polypeptide and its antagonists may be useful for
 CC treating or preventing disorders associated with the activity of
 CC LRSP. Such disorders include autoimmune/inflammatory disorders (for
 CC example AIDS, allergies, asthma, diabetes mellitus), bacterial and
 CC fungal infection and cancers (such as leukemia, adenocarcinoma,
 CC melanoma). Antibodies to LRSP may be useful for diagnosis of
 CC the above disorders.
 XX
 SQ Sequence 1331 BP; 397 A; 258 C; 228 G; 447 T; 1 other;
 Query Match 99.9%; Score 1330; DB 21; Length 1331;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTATGTTTTAGAAATCAAGATGAACCGGTAGCTGCTCATGTACCAACGTTGAATTT 60
 DB 1 CTATGTTTTAGAAATCAAGATGAACCGGTAGCTGCTCATGTACCAACGTTGAATTT 60
 QY 61 ACAGTGTTCACAAATGCTCGGAATTTTGCACCTGCCATAGGGAATGTTAAGTTTACTTGGC 120
 DB 61 ACAGTGTTCACAAATGCTCGGAATTTTGCACCTGCCATAGGGAATGTTAAGTTTACTTGGC 120
 QY 121 TGGAAATTTATCAGACTTGTGAGTAAACAAGTTTACAGATGAGGGGGAATTTATG 180
 DB 121 TGGAAATTTATCAGACTTGTGAGTAAACAAGTTTACAGATGAGGGGGAATTTATG 180
 QY 181 AGSCCCTTAGGCTTAACAAATAATCAGTATCTGAGATAGTGGCTAAATGTGCTCCCA 240
 DB 181 AGSCCCTTAGGCTTAACAAATAATCAGTATCTGAGATAGTGGCTAAATGTGCTCCCA 240
 QY 241 GGCCTAATTTGGGAACAGTTTTTCCCTGATTGCTTTGAGAAGTACTTTCTTTGACAGAAA 300
 DB 241 GGCCTAATTTGGGAACAGTTTTTCCCTGATTGCTTTGAGAAGTACTTTCTTTGACAGAAA 300
 QY 301 TTTTCAATTCGCTTGCCATGCTAATTCCTCTTTATAGGAGCCATGGAATTTCTTCC 360
 DB 301 TTTTCAATTCGCTTGCCATGCTAATTCCTCTTTATAGGAGCCATGGAATTTCTTCC 360
 QY 361 TTTTGTGGGAATGTCCTCAATAGCATTTTCAGATCTTTTGTGATGTCACCTAATGCCATTAT 420
 DB 361 TTTTGTGGGAATGTCCTCAATAGCATTTTCAGATCTTTTGTGATGTCACCTAATGCCATTAT 420
 QY 421 TGGTAATGCGGTATTGTTGTAATACAGCATAGTTAAATTAACCTGTTTACAGTAAATCTACA 480
 DB 421 TGGTAATGCGGTATTGTTGTAATACAGCATAGTTAAATTAACCTGTTTACAGTAAATCTACA 480
 QY 481 CTGGAATTTGCTGCACCTCTACCAATAGCCTTTTGAATGACCTGAAAGTCTTAACAGAGAA 540
 DB 481 CTGGAATTTGCTGCACCTCTACCAATAGCCTTTTGAATGACCTGAAAGTCTTAACAGAGAA 540
 QY 541 AGAGGATCTCTGCAGAAAGATAGCTAATATTTTGGTACTTTATCTGAAATCCAG 600
 DB 541 AGAGGATCTCTGCAGAAAGATAGCTAATATTTTGGTACTTTATCTGAAATCCAG 600
 QY 601 ATGCTGCTTCCCTGAGGTTGTTTCTTCTTACGATCCTCATGAAATCCCTCTGGGA 660

Db 601 ATGCTGCTTCCCTGAGGTTGTTTCTTCTTACGATCCTCATGAAATCCCTCTGGGA 660
 QY 661 GCACAGGACAGTATAGTAACCTCTCCATTTCTTTGTTTGTGTTTAAAGACAGAGACTCT 720
 Db 661 GCACAGGACAGTATAGTAACCTCTCCATTTCTTTGTTTGTGTTTAAAGACAGAGACTCT 720
 QY 721 GTCTCAAAAAAAGGACATTTATCATATATAACATCTTTATAGAGCCCTTAATTTCTTATC 780
 Db 721 GTCTCAAAAAAAGGACATTTATCATATATAACATCTTTATAGAGCCCTTAATTTCTTATC 780
 QY 781 TGAAGGCACCTGTTTTTTTTTAAACAGTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 840
 Db 781 TGAAGGCACCTGTTTTTTTTTAAACAGTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 840
 QY 841 GATCAGATAGTCCCTCTGTCACAGTAGCAAAATGCTGTTTCATATAAGTGGGAGAAACAG 900
 Db 841 GATCAGATAGTCCCTCTGTCACAGTAGCAAAATGCTGTTTCATATAAGTGGGAGAAACAG 900
 QY 901 CATTTTAAAGTAACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTT 960
 Db 901 CATTTTAAAGTAACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTT 960
 QY 961 AAGAAAAAACCCTTCCACCTTTACTGTGTCATTTATATCCCTTAGTTCACAAAGTTA 1020
 Db 961 AAGAAAAAACCCTTCCACCTTTACTGTGTCATTTATATCCCTTAGTTCACAAAGTTA 1020
 QY 1021 ATTATCTTATTTCTGATATTTTATACCAAGAGCTTTATCAGCCAGTTTCCAGAAC 1080
 Db 1021 ATTATCTTATTTCTGATATTTTATACCAAGAGCTTTATCAGCCAGTTTCCAGAAC 1080
 QY 1081 AACCCTATATACCAACCAATACCAACCAATACCAACCAATACCAACCAATATATATCATC 1140
 Db 1081 AACCCTATATACCAACCAATACCAACCAATACCAACCAATACCAACCAATATATATCATC 1140
 QY 1141 AGTAACCTGACGACATGATTTAGTGGCTTTGATTTGCAAAATACGACTTCTACATCCATAT 1200
 Db 1141 AGTAACCTGACGACATGATTTAGTGGCTTTGATTTGCAAAATACGACTTCTACATCCATAT 1200
 QY 1201 TCTCATCTTTCATACCATATACACTACTACCATCTTTTGTNAGATCATCTAAGAGCAAT 1260
 Db 1201 TCTCATCTTTCATACCATATACACTACTACCATCTTTTGTNAGATCATCTAAGAGCAAT 1260
 QY 1261 GCGAATGTAACCCCTATAATTTACTGGTACTCTTTGGTTCCAGATCTTGCCTTTCC 1320
 Db 1261 GCGAATGTAACCCCTATAATTTACTGGTACTCTTTGGTTCCAGATCTTGCCTTTCC 1320
 QY 1321 AATGTCACCTTG 1331
 Db 1321 AATGTCACCTTG 1331

RESULT 2
 AAA48964
 ID AAA48964 standard; DNA; 552 BP.
 XX
 AC AAA48964;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human statherin DNA.
 XX
 KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
 KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 73..261
 FT /*tag= a
 FT /product= Statherin

XX WO200024779-A1.
 XX 04-MAY-2000.
 XX 22-OCT-1999; 99WO-US24046.
 XX 23-OCT-1998; 98US-0155209.
 XX (INCY-) INCYTE PHARM INC.
 XX Tang YT, Corley NC, Guegler KJ, Patterson C;
 XX WPI; 2000-350699/30.
 XX P-PSDB; AAY94527.
 XX Purified polypeptide used for treating or preventing a disorder
 XX characterized by expression or activity of lysine-rich statherin
 XX proteins -
 XX
 XX Disclosure; Page 71; 75pp; English.
 XX
 XX The present invention relates to human lysine-rich statherin protein
 XX (LRSP)(AAY94526). The cDNA sequence encoding this protein was identified
 XX through analysis of a cDNA library of breast tumour tissue (BRSTN014).
 XX The LRSP sequence was found to have homology with the DNA of human
 XX statherin (the present sequence) and human basic histidine-rich protein
 XX (AAY94528). Human statherin is a phosphoprotein that acts as an
 XX inhibitor of precipitation of calcium phosphate salts in the oral cavity.
 XX The LRSP polypeptide and its antagonists may be useful for treating or
 XX preventing disorders associated with the activity of LRSP. Such
 XX disorders include autoimmune/inflammatory disorders (for example AIDS,
 XX allergies, asthma, diabetes mellitus), bacterial and fungal infection
 XX and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to
 XX LRSP may be useful for diagnosis of the above disorders.
 XX
 XX Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;

Query Match 19.0%; Score 253; DB 21; Length 552;
 Best Local Similarity 88.4%; Pred. No. 2e-53;
 Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 1022 TTATCTTATTCGGATATGCTTTTATACCAAGAGCCTTATCAGCCAGTTCAGAACCA 1081
 DB 150 TTGCGTAGAATGGGAAGATTCGGTTATGGGTATGGCTTATCAGCCAGTTCAGAACCA 209
 QY 1082 ACCACTATACGCACACCAATACCAACCAATACCAACCAATACCAATACCAATACCAAT 1141
 DB 210 ACCACTATACGCACACCAATACCAACCAATACCAACCAATACCAATACCAATACCAAT 269
 QY 1142 GTAACTGCAGGACATGATTATGAGGCTTGATGGCAATACGACTTCTACATCCATATT 1201
 DB 270 GTAACTGCAGGACATGATTATGAGGCTTGATGGCAATACGACTTCTACATCCATATT 329
 QY 1202 CTCATCTTTTCATACCATATCAGACTACTACCATCTTTTGTGAGATCATCATAGAGCAATG 1261
 DB 330 CTCATCTTTTCATACCATATCAGACTACTACCATCTTTTGTGAGATCATCATAGAGCAATG 389
 QY 1262 CGAATGAAACCCCTATATTTACTGATGACTCTTTGGTTCCAGATACCTTGTCTTCA 1321
 DB 390 CAATGAAACCACTATATTTACTGATGACTCTTTGGTTCCAGATACCTTGTCTTCA 449
 QY 1322 ATGTCACCTTG 1331
 DB 450 TTGTCACCTTG 459

RESULT 3
 AAC76665
 ID AAC76665 standard; CDNA; 2121 BP.
 XX
 AC AAC76665;
 XX

DT 08-FEB-2001 (first entry)
 XX Human ORFX ORF2220 polynucleotide sequence SEQ ID NO:4439.
 DE
 XX
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic; antithyroid;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disease; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 XX 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US08621.
 XX
 XX 31-MAR-1999; 99US-0127607.
 XX 02-APR-1999; 99US-0127636.
 XX 05-APR-1999; 99US-0127728.
 XX 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach M;
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB42456.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 XX useful for treating e.g. cancers, proliferative disorders,
 XX neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 5; Page 3631-3632; 5507pp; English.
 XX
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
 XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
 XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 XX antidiabetic; hypotensive; dermatological; immunosuppressive;
 XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 XX antithyroid; and antianaemic. The sequences can be used for determining
 XX the presence of or predisposition to, or preventing or treating
 XX pathological conditions associated with an ORFX-associated disorder. The
 XX nucleic acids can be used to express ORFX proteins in gene therapy
 XX vectors. The proteins and nucleic acids may be used to treat cancers,
 XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
 XX graft vs host disease, cardiovascular disease, diabetes mellitus,
 XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 XX coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 XX Sequence 2121 BP; 617 A; 430 C; 341 G; 732 T; 1 other;

Query Match 15.7%; Score 209.6; DB 21; Length 2121;
 Best Local Similarity 81.5%; Pred. No. 2.2e-42;
 Matches 255; Conservative 0; Mismatches 55; Indels 3; Gaps 1;

QY 1022 TTATCTTATTTTGGATATGCTTTTATACCAAGAGCCCTTATCAGCCAGTTCCAGAAC 1081
 Db 1578 TTGGCGTAGAATGGGAAGATTCGGTTATGGGTATGCGCCCTTATCAGCCAGTTCCAGAAC 1637
 QY 1082 ACCACTATACGCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1141
 Db 1638 ACCACTATACGCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1697
 QY 1142 GTAACGAGGACATGATTATGAGGCTTGTGAGTGGCAATACGACCTTCTACATCCATAT 1201
 Db 1698 GTAACGAGGACATGATTATGAGGCTTGTGAGTGGCAATACGACCTTCTACATCCATAT 1757
 QY 1202 CTCATCTTTTCATACCATACACATACACATACACATTTTGTGNAGATCACTACAGCAATG 1261
 Db 1758 CTCATCTTTTCATACCATACACATACACATTTTGTGAAGAAATATACATAAGCAATG 1817
 QY 1262 CGAATGTA---AAACCCCTATATTTTACTGGATACCTTTTGGTCCAGATACCTTGGCTTTT 1318
 Db 1818 CAGATAAAGAAATACCATGATTACTGATACCTTTTGGTCCAGATACCTTGGCTTTT 1877
 QY 1319 CCAATGTCACATG 1331
 Db 1878 TAATTATCATTTG 1890

RESULT 4

AA42355
 ID AA42355 standard; cDNA; 221 BP.

AC AA42355;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1095.

XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
 KW antitumor; osteoprotective; neuroprotective; nontropic; antiproliferative;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune diabetes; asthma; myeloid cell deficiency; ulcer;
 KW insulin dependent diabetes; multiple sclerosis; allergic condition;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021990-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24205.

XX 15-OCT-1998; 98US-0104435.

XX (GEMV) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;

DR WPI; 2000-317937/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -

XX Claim 1; Page 394; 618pp; English.

XX

CC AA411261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiasthmatic; antitumor; antiparkinsonian; osteoprotective;
 CC nontropic; antiproliferative; neuroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention.

XX SQ Sequence 221 BP; 72 A; 41 C; 44 G; 64 T; 0 other;

Query Match 14.1%; Score 187.4; DB 21; Length 221;
 Best Local Similarity 99.5%; Pred. No. 3.6e-37;
 Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 793 TTTTNTTTTAAACAGTTAAGTACTGATGTCACACAGACAATATTTCTGATCAGATAGTC 852
 Db 28 TTTTNTTTTAAACAGTTAAGTACTGATGTCACACAGACAATATTTCTGATCAGATAGTC 87
 QY 853 CCCTGTCAACAGTAGCAAAATGTGGTTTCATAAAGTGGGAAGAAACAGCATTTTAAAGTA 912
 Db 88 CCCTGTCAACAGTAGCAAAATGTGGTTTCATAAAGTGGGAAGAAACAGCATTTTAAAGTA 147
 QY 913 ACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTTAAGAAAAA 972
 Db 148 ACTTTTGGGAGACTGATTTGAGTAATAATAAACTCTGGTCTCCCTTAAGAAAAA 207
 QY 973 CCCTTCCAC 981
 Db 208 CCCTTCCGC 216

RESULT 5

AAI91635

ID AAI91635 standard; cDNA; 390 BP.

XX AAI91635;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 11695.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX


```

XX 11-NOV-1994; 94WO-JP01916.
XX 12-NOV-1993; 93JP-0355504.
XX (MATSU/) MATSUBARA K.
XX (OKUBO/) OKUBO K.
XX Matsubara K, Okubo K;
XX WPI; 1995-206931/27.
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX for diagnosis of abnormal cell function, by preparing cDNA that
XX reflects relative abundance of corresp. mRNA in specific human
XX tissues
XX Claim 1; Page 1297; 2245pp; Japanese.
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX given in AAT19001-T26837 and which is able to hybridise to part of
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX sequences were obtained from 3'-directed cDNA libraries prepared
XX from various human tissues; synthesis of cDNA was initiated from the
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX untranslated sequence is unique to a particular mRNA species, almost
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX is constructed so as to reflect accurately the relative abundance of
XX different mRNAs in the particular tissue from which it was derived.
XX The appearance frequency of a given GS in a cDNA library can be
XX determined (esp. using primers and probes derived from the GS
XX sequences) as a means of diagnosing abnormal cell function or for
XX recognising different cell types.
XX Sequence 90 BP; 30 A; 12 C; 17 G; 31 T; 0 other;
XX
Query Match 5.7%; Score 76.4; DB 16; Length 90;
Best Local Similarity 97.8%; Pred. No. 1.4e-09;
Matches 88; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
XX
OY 392 GATCTTTTCATGTGCCTAATGCCATTATTTGGTAATGCC-GTTATTTGGTAATACAGCAT 450
DB 1 GATCTTTTCATGTGCCTAATGCCATTATTTGGTAATGCCGTTATTTGGTAATACAGCAT 60
OY 451 ACTTAATAAAGTGTACAGTAATCTACA 480
DB 61 AGTTAATAAAGTGTACAGTAATCTACA 90
XX
RESULT 10
ABL54353
ID ABL54353 standard; DNA; 9881 BP.
XX
AC ABL54353;
XX
DT 29-JUL-2002 (first entry)
XX
DE Chemically treated apoptosis gene #27.
XX
KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
KW amyotrophic lateral sclerosis; cancer; ds.
XX
OS Unidentified.
XX
PN WO200177164-A2.
XX
PD 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP03969.
XX
PR 06-APR-2000; 2000DE-1019058.

```

```

PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017444/02.
XX
XX Chemically modified sequences of genes associated with apoptosis are
XX useful to determine methylation patterns of genomic DNA samples for
XX diagnosis of associated diseases such as cancer.
XX Claim 1; Seq ID #53; 24pp; English.
XX
XX This invention relates to chemically pre-treated DNA of genes
XX associated with apoptosis. The nucleic acids are used to allocate
XX patients for specific therapy for HIV infection, Bloom syndrome,
XX cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
XX infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
XX and cancers. This nucleotide sequence represents a chemically
XX treated apoptosis gene. Even SEQ ID numbers are the complementary
XX DNA strands to the odd SEQ ID numbers. The sequence data for this
XX patent is not represented in the printed specification but is based on
XX information supplied by the European patent office.
XX
XX Sequence 9881 BP; 2606 A; 196 C; 2308 G; 4771 T; 0 other;
XX
Query Match 5.6%; Score 74.6; DB 24; Length 9881;
Best Local Similarity 78.8%; Pred. No. 2e-08;
Matches 89; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
XX
OY 860 AACAGTAGCAATGTGGTTTCATAGAGTGGGAAGAAACAGCATTTTAAAGTAACCTTTT 919
DB 5771 AAGAGTAGTAATGTGGTTTATTAAGTGGGAAGAAAGTAGTAAATTTAAATAATTTT 5830
OY 920 GGGAGACGATTTTGGTAATATAATAAACTCTGCTCCCTTAAGAAAAA 972
DB 5831 GGGAGATTAATGAGTAATAATAAAATTTTGTGTTTTCGTTAATAATAA 5883
XX
RESULT 11
AAV89525
ID AAV89525 standard; cDNA; 516 BP.
XX
AC AAV89525;
XX
DT 15-FEB-1999 (first entry)
XX
DE EST clone CP289.
XX
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
XX tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic;
XX receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
XX gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO9845436-A2.
XX
PD 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US06955.
XX
XX 10-APR-1997; 97US-0838821.
XX
XX (GEMV) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX Racie LA, Spaulding V, Treacy M;
XX WPI; 1999-070077/06.

```

XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1: Page 245; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other;

Query Match 4.5%; Score 60.2; DB 20; Length 516;
Best Local Similarity 89.0%; Pred. No. 2.9e-05;
Matches 65; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1047 TATACCAAGAGCGCTATCAGCCAGTTCAGAACACCACTATACGACACCAATACCAA 1106
DB 139 TATGGGTATGGCCCTTATCAGCCAGTTCAGAACACCACTATACGACACCAATACCAA 198
QY 1107 CCACATATACCAAC 1119
DB 199 CCACATATACCAAC 211

RESULT 12
ABL43273
ID ABL43273 standard; DNA; 60 BP.
AC ABL43273;
XX
XX 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:16021.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
XX 20-JUL-2001; 2001WO-IB01903.
XX
XX 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
XX New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes.
XX
XX Example 1; SEQ ID 16021; 47pp; English.
PS
XX The present invention describes oligonucleotide libraries for detecting

CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABL27253 to ABL59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 60 BP; 22 A; 25 C; 4 G; 9 T; 0 other;

Query Match 4.2%; Score 56.4; DB 24; Length 60;
Best Local Similarity 98.3%; Pred. No. 0.00012;
Matches 57; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1059 CCTATCAGCCAGTTCAGACACCACTATACGACACCAATACCAATACC 1116
DB 3 CCTTATCAGCCAGTTCAGACACCACTATACGACACCAATACCAATACC 60

RESULT 13
ABL54354/C
ID ABL54354 standard; DNA; 9881 BP.
XX
XX ABL54354;
XX
XX 29-JUL-2002 (first entry)
XX
DE Chemically treated apoptosis gene complementary to gene #27.
XX
XX Apoptosis; HIV; Bloom syndrome; cardiopathy;
KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
KW amyotrophic lateral sclerosis; cancer; ds.
XX
OS Unidentified.
XX
XX WO200177164-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP03969.
PF
XX 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-017444/02.
DR
XX Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer.
XX
XX Claim 1; Seq ID #54; 24pp; English.

XX This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented in the printed specification but is based on
 CC information supplied by the European patent office.

XX Sequence 9881 BP; 2874 A; 196 C; 2093 G; 4718 T; 0 other;
 Query Match 4.2%; Score 55.8; DB 24; Length 9881;
 Best Local Similarity 67.1%; Pred. No. 0.001;
 Matches 96; Conservative 0; Mismatches 42; Indels 5; Gaps 1;
 QY 830 CAATATTTCTGATCAGATAGTCCCTGTCACAGTAGCAAAATGTTTCATATAAGTGG 889
 DB 4136 CAATATTTCTGATCAGATAGTCCCTGTCACAGTAGCAAAATGTTTCATATAAGTGG 889
 QY 890 GAAGAAACAGCAGATTTTAAAGTAACTTTTGGGAGACTGATTTGACTAATAATAAACTC 949
 DB 4081 AAAAAAACAACAATTTAAATFARCTTTTAAAAAACTAAATTAATAATAATAAACTT 4022
 QY 950 TGCTCTCCCTTAGAAAAAATA 972
 DB 4021 CAATCTTTGCTAATAATAATA 3999

RESULT 14
 ABL33026
 ID ABL33026 standard; DNA; 6056 BP.
 AC ABL33026;
 XX 26-MAR-2002 (first entry)
 XX Human immune system associated gene SEQ ID NO: 999.
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosine; neotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO2000200928-A2.
 XX 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX Claim 1; SEQ ID NO 999; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 6056 BP; 1780 A; 54 C; 1079 G; 3143 T; 0 other;

Query Match 3.6%; Score 48.4; DB 24; Length 6056;
 Best Local Similarity 45.9%; Pred. No. 0.06;
 Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 688 TTCTTTTGGTTTGTGTTTAAAGACAGAGACTCTGCTCAAAAAAGGACATTTATCATT 747
 DB 3741 TTTTATGATTTTCGAAATGAAAAATTTTATTTGTTTAAATAAATAAATAAATAA 3800
 QY 748 ATAACATCTTATAGAGCCCCCTAATTTCTATCTGAAGGACACGTGTTTTTAAACA 807
 DB 3801 AGTAGGTTTATGTAAGTATTTTGTGTTTGTATTTTAGGAATTTTGTGTTTATATA 3860
 QY 808 GTTAAGTACTGATGTCACAGACAAATATTTCTGATCAGATAGTCCCTGTCACAGTAG 867
 DB 3861 ATTTTGTATGTTAAATATAAATAAATAAATGTTATTTGTTTAAATTTTATAGTAA 3920
 QY 868 CAAATGTTGTTTCAATAAGTGGGAAGAAACAGACATTTTAAAGTAACTTTTGGGAGACT 927
 DB 3921 GTAATATAGTTAGTATATGTTTAAAGAAAAAATGTTAAAGAAATTTCTTTTATAT 3980
 QY 928 GATTGAGCTAATAATAAACTCTGGCTCCCTTAAGAAAAAACCCCTTCCACCTTTAC 987
 DB 3981 TATTTTAAAGATGTTGATATTTATTTGTTTAAACGTTATATTTTATAAGATTAT 4040
 QY 988 TGTGTCATTTATATCCCTTAGTTCCAAAGTTAATTTATCTTATTTCTGGATATGCTTTT 1047
 DB 4041 TTTTAAATGTTGAATATAAGTTTAAATATATATGTTTAAATTTGTTTAAATAT 4100
 QY 1048 AT 1049
 DB 4101 TT 4102

RESULT 15
 AAS46787/C
 ID AAS46787 standard; DNA; 61020 BP.
 XX AAS46787;
 AC AAS46787;
 XX 18-DEC-2001 (first entry)
 XX Tumour suppressor gene derived chemically modified sequence #513.
 DE Human; tumour suppressor gene; oncogene; antitumour; cytosine;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO200168912-A2.
 XX 20-SEP-2001.
 XX 15-MAR-2001; 2001WO-EP02955.
 XX 15-MAR-2000; 2000DE-1013847.
 XX 06-APR-2000; 2000DE-1019058.
 XX 07-APR-2000; 2000DE-1019173.
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.

Db 216 NCTGCANCAANTGATTTTACACCTTACTAGAAACAACTAANGAAAGCACTNATTAGCT 275
QY 215 GAGATAGTGGCTAATGTGGCTCCCGAGGCTAATTTGGGAACAGTTTTTC 264
Db 276 CTGAATNAANTNACATGGNAAGCCTTTTACTAATCTNCAAAANAAACCTTC 325

RESULT 4
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5385526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kuritz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo

REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:

NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:

NAME/KEY: CDS
; LOCATION: 2378..5038
US-07-867-106-2

Query Match 2.8%; Score 37.2; DB 1; Length 5852;
Best Local Similarity 52.6%; Pred. No. 2.2;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 682 TCTCCATTTCTTTGTTTCTTTTAAACAGACAGACTGTCTCAAAAAAGGACATTT 741
Db 1973 TTTCTATTCTTTCTTTTCTTTTAACTTTGAAAAAAGGACATTT 741

QY 742 ATCATATAACATCTTATTAGAGCCCTAATTTCTATCTGAAGCAGCTGTTTTTTT 801
Db 2033 AAAAAAACCCTCATTAATAATTAATTAATTTGTTTTTTTGAATTTTTTTTA 2092

QY 802 TAACAGTTAAGTACTGATGTCACAGACAAAATA 835

Db 2093 ATAAATTAAATTTTATCTCTATCTAATATA 2126

RESULT 5

US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996

CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs
; TYPE: nucleic acid

STRANDEDNESS: not relevant
; TOPOLOGY: not relevant

MOLECULE TYPE: cDNA
; FEATURE:

NAME/KEY: misc.feature
; LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 2.8%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 84.0%; Pred. No. 8.5;
Matches 42; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 912 AACTTTTGGGAGACTGATTGAGTAATAATAAACTGGTCTCCCTTA 961
Db 154422 AATTCTCAAGGAGACTGATTGAGTAATAATAAACTGCTCTCTGTA 154471

RESULT 6

US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.

; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..12730
; OTHER INFORMATION: /note= "RG2C"
US-09-004-838-91

Query Match 2.7%; Score 36.4; DB 4; Length 12730;
Best Local Similarity 51.2%; Pred. No. 4.7;
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 678 GAACCTCCATTCCTTTGTTGTTTAAAGACAGAGACTCTGTCTCAAAAAAGGAC 737
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 3997 GAATGACATTTAAAGTTGTTAACTTATAAATCCGTTGCTTATTTCACATAAAGTAATT 3938
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 738 ATTATCATTAACATCTATTAGAGCCCTAAATTTCTTAATCTGAAGGCACTGTTTTT 797
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3937 ATTACATTTAAGCATTTATTAGCTAAAGTTAGTAATTAGTTCTTAATTTAAATGCTTT 3878
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 798 TTTTAAACAGTTAAGTACTGTCTCAACAGACAAATATTTCTGAT 843
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3877 AAACCTACTATTATAATGATCTCTAATGAATAATCATGTTGAT 3832
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-09-004-838-124/c
; Sequence 124, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Michelmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; TITLE OF INVENTION: Confering pest Resistance in Plants
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-078810US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 124:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12793 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..12793
; OTHER INFORMATION: /note= "RG2S"
US-09-004-838-124

Query Match 2.7%; Score 36.4; DB 4; Length 12793;
Best Local Similarity 51.2%; Pred. No. 4.7;
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 678 GAACCTCCATTCCTTTGTTGTTTAAAGACAGAGACTCTGTCTCAAAAAAGGAC 737
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 3849 GAATGACATTTAAAGTTGTTAACTTATAAATCCGTTGCTTATTTCACATAAAGTAATT 3790
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 738 ATTATCATTAACATCTATTAGAGCCCTAAATTTCTTAATCTGAAGGCACTGTTTTT 797
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3789 ATTACATTTAAGCATTTATTAGCTAAAGTTAGTAATTAGTTCTTAATTTAAATGCTTT 3730
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 798 TTTTAAACAGTTAAGTACTGTCTCAACAGACAAATATTTCTGAT 843
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3729 AAACCTACTATTATAATGATCTCTAATGAATAATCATGTTGAT 3684
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-09-134-001C-748/c
; Sequence 748, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 748
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-748

Query Match 2.7%; Score 36.2; DB 4; Length 762;
Best Local Similarity 54.0%; Pred. No. 1.9;
Matches 74; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 586 CATTCTCTTTGTTGTTTAAAGACAGAGACTCTGTCTCAAAAAAGGACATTATCA 745
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 181 CGTTATTGAAATAATTTTACATTTGAATACATTTTCATCTCGAAAAAGTGACAATAAT 122
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 746 TTATAACATCTTATTAGAGCCCTAAATTTCTTAATCTGAAGGCACTGTTTTTAAAA 805
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 TTGTCTCTTCTTATAGCGCATACATTACGATATGATAGTCTCTTTGTTTCAA 62
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 806 CAGTTAAGTACTGATGT 822
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TAAAAAGCACTGATTT 45
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-234-827B-1/c
; Sequence 1, Application US/09234827B
; Patent No. 6448471
; GENERAL INFORMATION:
; APPLICANT: Fuzio, Piotr S.
; APPLICANT: Grudler, Florian M.W.
; TITLE OF INVENTION: Nematode feeding structure specific gene and its

;; TITLE OF INVENTION: application to produce nematode resistant plants
;; FILE REFERENCE: U-012084-2
;; CURRENT APPLICATION NUMBER: US/09/234,827B
;; CURRENT FILING DATE: 1999-01-21
;; PRIOR APPLICATION NUMBER: US 60/072,142
;; PRIOR FILING DATE: 1998-01-22
;; NUMBER OF SEQ ID NOS: 32
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 8302
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (4038)..(4256)
;; NAME/KEY: CDS
;; LOCATION: (4807)..(5604)
;; NAME/KEY: CDS
;; LOCATION: (6777)..(6827)
;; NAME/KEY: CDS
;; LOCATION: (6919)..(7935)
US-09-234-827B-1

Query Match 2.7%; Score 36.2; DB 4; Length 8302;
Best Local Similarity 51.2%; Pred. No. 4.5;
Matches 111; Conservative 0; Mismatches 103; Indels 3; Gaps 1;
QY 683 CTCACATTCCTGTTGTTTAAACAGACAGACTGCTCTCAAAAAAGACATTTA 742
Db 805 CTTATGTTGTTGATTTTTTATAGACAGTCCAGCGAAACACATCAATACACTTA 746
QY 743 TCATTATAACATCTATTAGAGCCCAATTTCTTATCTGAAGCAGCTGTTTTTTTT 802
Db 745 ATTTTAAATCATATATACATTTTAAATTCGTGTAA---TTATTTCTAAATTA 689
QY 803 AAACAGTTAGTACTGATGCAACAGACAAATATTTCTGATCAGATGCCCCGTGCAAC 862
Db 688 AATAATTAATATTAATATGTCCTCAATATAAATGATGATGTTTATGGGTTTACC 629
QY 863 AGTAGCAATGTTGTTTCATTAAGTGGGAAGAAACA 899
Db 628 TGTAAGCACTCTGTTTATCAAAAAAGAAAAAAA 592

RESULT 12
US-09-424-283-6
;; Sequence 6, Application US/09424283
;; Patent No. 6437219
;; GENERAL INFORMATION:
;; APPLICANT: Grimes, et al.
;; TITLE OF INVENTION: Sucrose binding proteins
;; FILE REFERENCE: 4630-50206
;; CURRENT APPLICATION NUMBER: US/09/424,283
;; CURRENT FILING DATE: 1999-11-19
;; PRIOR APPLICATION NUMBER: PCT/US98/10465
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/047,568
;; PRIOR FILING DATE: 1997-05-22
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 6
;; LENGTH: 3718
;; TYPE: DNA
;; ORGANISM: Glycine max
US-09-424-283-6

Query Match 2.7%; Score 36; DB 4; Length 3718;
Best Local Similarity 47.0%; Pred. No. 3.8;
Matches 111; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 983 TTTACTGTCTATATATCCCTTAGTTCACAAAGTTAATTAATCTTATTTCTGGATATG 1042
Db 3257 TGTACGCATTAATTGATTTTTTTCTCCACATTAATAGGTGAATCAGTTAGAGAA 3316

QY 1043 CTTTATACCAAGACGCTTATCAGCCAGTTCACGAACACACATATACGCACACCATTA 1102
Db 3317 ATATTAATAAATAAATAAATAAAGGAGAGAGCTAATACAAATAAGAAATACGAATC 3376
QY 1103 CCAACACAAATACCAACATATATACGTTTAAATATATCATCATCAGTAACCTGCAGGACATGATTAT 1162
Db 3377 ACAATGAATAGACCAATTAGAACCATTTATTTCTTACAAATTTAAAGAAAAAGCTTTT 3436
QY 1163 TGAGGCTTGATGGCAAAATACGACTTCACATCCATATATCTCAATCTTTCATACCAT 1218
Db 3437 TTAACAATATATACATTTATCATCTATATATATTTTATTTATATATTTTATAACTT 3492

RESULT 13
US-09-009-913-1
;; Sequence 1, Application US/09009913
;; Patent No. 6087485
;; GENERAL INFORMATION:
;; APPLICANT: Axys Pharmaceuticals, Inc.
;; TITLE OF INVENTION: Asthma Related Genes
;; NUMBER OF SEQUENCES: 339
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Bozicevic & Reed, LLP
;; STREET: 285 Hamilton Ave, Suite 200
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94301
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/009,913
;; FILING DATE: 21-JAN-1998
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sherwood, Pamela J
;; REGISTRATION NUMBER: 36,677
;; REFERENCE/DOCKET NUMBER: SEQ-4P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-327-3231
;; TELEFAX: 650-327-3231
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 72928 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 2.7%; Score 35.8; DB 3; Length 72928;
Best Local Similarity 54.1%; Pred. No. 13;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 708 AGACAGAGACTCTGCTCAAAAAAAGGACATTTATCATTAATCAATCTTTATAGAGCC 767
Db 68237 AGAGTGAGACTCTGCTCAAAAAAATTAATAAAAAAACTCTGTAATTTACTTTG 68296
QY 768 CTAATTTCTGTAAGGACACTGTTTTTTTTTAAACAGTTAACTACTGATGCAACA 827
Db 68297 CACCACATAATATGATATCACACATTTATTTAAANAATATTTTACATTTGCTTTTAA 68356
QY 828 GACAAATATTTCTGA 842
Db 68357 TATAAATTTTTTAA 68371

RESULT 14
US-09-134-001C-920/c
; Sequence 920, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 920
; LENGTH: 3780
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-920

Query Match 2.7%; Score 35.6; DB 4; Length 3780;
Best Local Similarity 50.8%; Pred. No. 4.9;
Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 273 TTGAGAGTACTTTCTTTTGACAGAAATTTTCATTCTGCTTCCCATTTGCTATATCTCC 332
Db 3049 TTTCAGAGCAGCTCTTTAGAAATATCTTTAGTTCATACCTGTTGATTTCTAAAT 2990
QY 333 CTTTATAGGAGCATTGGATTCTTTCTCTTTGCGAAATGCCATTAGCATTTTCAG 392
Db 2989 CTTTATAGTCTTTGTAATTTGCGTTTGTAGCAGAGGATATATTCAAATGCAT 2930
QY 393 ATCTTTGTATGTCACATAATGCCATTATGCTAATGCCGTTATGTTGAA 442
Db 2929 CTGCTACTTTCGCTGATCTATCTCTCGGCTATCATCTTTTAACTGAA 2880

RESULT 15
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5933827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned

REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13
Query Match 2.7%; Score 35.6; DB 2; Length 19124;
Best Local Similarity 45.0%; Pred. No. 8.8;
Matches 134; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 903 TTTTAAAGTAACCTTTTGGGAGACTGATTGTGAGTAATAATAAACTCTGTCTCCCTTAA 962
Db 6981 TTTTAAAAAAT 6922
QY 963 GAAAAAACCCTTCCACCTTTTACTGTGTCATTATATCCCTTAGTCCCAAGTTAAT 1022
Db 6921 AAAAAAAT 6862
QY 1023 TATCTTATTTCTGGATATTTGCTTTTATACCAAGAGCCTTATCAGCCAGTTCCAGACAA 1082
Db 6861 TATATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 6802
QY 1083 CCACATATAGCACAACCATACCAACCAATACCAACATATACGTTTAAATATCATCAG 1142
Db 6801 AATACGAAATCAACAT 6742
QY 1143 TAATGTCAGGACATGATTATTTGAGGCTTGTGCTTGGCAATACGACTTCTACATCATAT 1200
Db 6741 TAACATATATATATATATATATTTTGTGACTTTTATTTTATTTATATATATATAT 6684

Search completed: April 14, 2003, 23:07:09
Job time : 809 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 21:17:40 ; Search time 134 Seconds
(without alignments)
8712.764 Million cell updates/sec

Title: US-09-830-244B-2
Sequence: 1 ctatgttttagaatacaag.....gccttttccaatgctcacttg 1331

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues 1186858

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications, NA.*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
 - 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
 - 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
 - 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
 - 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104.6	7.9	438	9	US-09-924-340-5
2	104.6	7.9	438	9	US-09-924-340-5
3	74.8	5.6	433	10	US-09-983-965-5821
4	46.8	3.5	3104	9	US-10-037-598-32
5	46.8	3.5	513509	9	US-09-754-853A-4
6	43.2	3.2	65359	10	US-09-804-472-3
7	42.2	3.2	6055	9	US-10-091-504-1746
8	42.2	3.2	6055	10	US-09-764-869-1746
9	42.2	3.2	7809	9	US-10-091-504-1744
10	42.2	3.2	7809	9	US-10-091-504-1745
11	42.2	3.2	7809	10	US-09-764-869-1744
12	42.2	3.2	7809	10	US-09-764-869-1745
13	40.6	3.1	556	9	US-10-092-154-1685
14	40.6	3.1	556	9	US-09-764-847-1685
15	40.4	3.0	17849	9	US-10-092-154-1315
16	40.4	3.0	17849	10	US-09-764-847-1315
17	40.4	3.0	17862	9	US-10-092-154-1313
18	40.4	3.0	17862	10	US-09-764-847-1313
19	39.6	3.0	151	10	US-09-864-761-22646

20	39.6	3.0	492	10	US-09-864-761-5885
21	39.6	3.0	2000	9	US-09-938-842A-4305
22	39.6	3.0	44848	9	US-09-988-113-42
23	39.6	3.0	44848	10	US-09-776-874A-42
24	39	2.9	378361	9	US-09-901-136-3
25	38.6	2.9	5236	10	US-09-925-297-363
26	38.6	2.9	6265	10	US-09-129-112-3
27	38.6	2.9	10758	12	US-10-044-090-61
28	38.2	2.9	2000	9	US-09-938-842A-3631
29	38.2	2.9	2286	9	US-10-228-796-3
30	38.2	2.9	2286	10	US-09-191-687B-3
31	38	2.9	1282	9	US-09-938-842A-4177
32	38	2.9	33023	10	US-09-880-107-3350
33	37.4	2.8	13715	7	US-08-781-986A-195
34	37.4	2.8	397658	10	US-09-813-320-3
35	37.2	2.8	4000	9	US-09-938-842A-2721
36	37	2.8	2000	9	US-09-764-872-517
37	36.8	2.8	32204	9	US-10-072-349-327
38	36.8	2.8	32204	10	US-09-764-855-327
39	36.8	2.7	495	9	US-10-092-154-219
40	36.6	2.7	495	10	US-09-764-847-219
41	36.6	2.7	1009	9	US-10-076-816-18
42	36.6	2.7	2000	9	US-09-938-842A-2839
43	36.6	2.7	99014	10	US-09-880-107-3428
44	36.4	2.7	376	10	US-09-960-352-9302
45	36.2	2.7			

ALIGNMENTS

RESULT 1

US-09-924-340-5

; Sequence 5, Application US/09924340

; Publication No. US20030027248A1

; GENERAL INFORMATION:

; APPLICANT: Bejanin, Stephanie

; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

; FILE REFERENCE: 91.US2.REG

; CURRENT APPLICATION NUMBER: US/09/924,340

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574

; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 112

; SOFTWARE: JPatent

; SEQ ID NO 5

; LENGTH: 438

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: 5'UTR

; LOCATION: 1..83

; NAME/KEY: CDS

; LOCATION: 84..317

; NAME/KEY: 3'UTR

; LOCATION: 318..438

; NAME/KEY: polyA_signal

; LOCATION: 397..402

; NAME/KEY: polyA_site

; LOCATION: 423..438

US-09-924-340-5

Query Match 7.9%; Score 104.6; DB 9; Length 438;

Best Local Similarity 88.3%; Pred. No. 4.6e-17;

Matches 113; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 2276 AAGAGAGGAGTACTTTCTAGTTAATGAATGAATGTAACCGTACTCAGGAAATGTTTTAA 2335
QY 997 TATATCCCTTAGTTCCTCAAGTTAATATCTTATTTCTGGATATGCTTTTATACCAAG 1056
Db 2336 AAGACTGCCAGACTCTTTTAAAGCTTTAACTATGTTATTTCTGCTTTTATCTAACT 2395
QY 1057 AGCCTTATCAGCCAGTTCAGACACCACTATACGCAACAACATACCAACCAATACC 1116
Db 2396 ATAGGAATATGATTAATCTAGTGAAGAGATGAATTAATAGTTATTTGACATTTT 2455
QY 1117 AACATATATAGTTTAAATATCATCACTAAGTGCAGGACATGATTTAGGCGTTGATGG 1176
Db 2456 AAATAAATATGATTTTCTTATATGATATAAATAAAGACCTGTTATATGATTTTAA 2515
QY 1177 CAATACGACTTCTACATCCATATCTCATCTTTTATACCATATCACACTACTACCCTT 1236
Db 2516 GTAAACAACACTTTTAAAGTACAGCCTTATATTTAAATCTGTAGGCTAGATCAGAAT 2575
QY 1237 TTTGTNAGATCA 1248
Db 2576 AGTGCCATAGCA 2587

RESULT 11
US-09-764-869-1744
; Sequence 1744, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1744
; LENGTH: 7809
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1744

Query Match 3.2%; Score 42.2; DB 10; Length 7809;
Best Local Similarity 45.8%; Pred. No. 2.1;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 937 AATAATAAAACTCTGGTCTCCCTTAAGAAAAAACCCTTCCACCTTTACTGTGTCATT 996
Db 2277 AAGAGAGGAGTACTTTCTAGTTAATGAATGTAACCGTACTCAGGAAATGTTTTAA 2336
QY 997 TATATCCCTTAGTTCCTCAAGTTAATATCTTATTTCTGGATATGCTTTTATACCAAG 1056
Db 2337 AAGACTGCCAGACTCTTTTAAAGCTTTAACTATGTTATTTCTGCTTTTATCTAACT 2396
QY 1057 AGCCTTATCAGCCAGTTCAGACACCACTATACGCAACAACATACCAACCAATACC 1116
Db 2397 ATAGGAATATGATTAATCTAGTGAAGAGATGAATTAATAGTTATTTGACATTTT 2456
QY 1117 AACATATATAGTTTAAATATCATCACTAAGTGCAGGACATGATTTAGGCGTTGATGG 1176
Db 2457 AAATAAATATGATTTTCTTATATGATATAAATAAAGACCTGTTATATGATTTTAA 2516
QY 1177 CAATACGACTTCTACATCCATATCTCATCTTTTATACCATATCACACTACTACCCTT 1236
Db 2517 GTAAACAACACTTTTAAAGTACAGCCTTATATTTAAATCTGTAGGCTAGATCAGAAT 2576
QY 1237 TTTGTNAGATCA 1248
Db 2577 AGTGCCATAGCA 2588

RESULT 12
US-09-764-869-1745

; Sequence 1745, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1745
; LENGTH: 7809
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1745

Query Match 3.2%; Score 42.2; DB 10; Length 7809;
Best Local Similarity 45.8%; Pred. No. 2.1;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 937 AATAATAAAACTCTGGTCTCCCTTAAGAAAAAACCCTTCCACCTTTACTGTGTCATT 996
Db 2276 AAGAGAGGAGTACTTTTCTAGTTAATGAATGTAACCGTACTCAGGAAATGTTTTAA 2335
QY 997 TATATCCCTTAGTTCCTCAAGTTAATATCTTATTTCTGGATATGCTTTTATACCAAG 1056
Db 2336 AAGACTGCCAGACTCTTTTAAAGCTTTAACTATGTTATTTCTGCTTTTATCTAACT 2395
QY 1057 AGCCTTATCAGCCAGTTCAGACACCACTATACGCAACAACATACCAACCAATACC 1116
Db 2396 ATAGGAATATGATTAATCTAGTGAAGAGATGAATTAATAGTTATTTGACATTTT 2455
QY 1117 AACATATATAGTTTAAATATCATCACTAAGTGCAGGACATGATTTAGGCGTTGATGG 1176
Db 2456 AAATAAATATGATTTTCTTATATGATATAAATAAAGACCTGTTATATGATTTTAA 2515
QY 1177 CAATACGACTTCTACATCCATATCTCATCTTTTATACCATATCACACTACTACCCTT 1236
Db 2516 GTAAACAACACTTTTAAAGTACAGCCTTATATTTAAATCTGTAGGCTAGATCAGAAT 2575
QY 1237 TTTGTNAGATCA 1248
Db 2576 AGTGCCATAGCA 2587

RESULT 13
US-10-092-154-1685
; Sequence 1685, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1685
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1685

Query Match 3.1%; Score 40.6; DB 9; Length 556;
Best Local Similarity 55.2%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 575 TTTTGTACTTTATCTGAAATCCAAGATGCTGCTCCCTGCGAGGTGTTTCTTCTTCTTA 634
Db 88 TTTAAAAAATTTTTTTTAAATGCACCTCTTCTCCCGTCTGGATATATCATCTACT 147

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2003, 21:02:56 ; Search time 1828 Seconds
(without alignments)
11792.228 Million cell updates/sec

Title: US-09-830-244B-2
Perfect score: 1331
Sequence: 1 ctatgttttttagaatcaag.....gccttttccaatgtaacttg 1331

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mus.*
24: em_gss_mam.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	490	36.8	492	9 AI224097	AI224097 q135a03.x
C 2	480	36.1	518	9 AI694371	AI694371 wd83a04.x
C 3	478.4	35.9	523	10 AW269813	AW269813 xv45f02.x
C 4	405	30.4	888	12 BG121534	BG121534 602352872
C 5	399	30.0	427	12 BF831772	BF831772 PM3-HT090
C 6	398.4	29.9	417	12 BE767736	BE767736 RC3-GN004

C 7	380.4	28.6	394	12 BG011356	BG011356 RC1-GN026
C 8	369.4	27.8	403	12 BF831960	BF831960 PM3-HT090
C 9	353	26.5	408	12 BG011354	BG011354 RC3-GN026
C 10	352.8	26.5	392	12 BE929791	BE929791 RC3-GN004
C 11	345.6	26.0	402	12 BG011351	BG011351 RC1-GN026
C 12	344.8	25.9	395	12 BF996088	BF996088 RC3-GN004
C 13	335.2	25.2	951	12 BG168510	BG168510 602333663
C 14	333	25.0	396	12 BF832034	BF832034 PM3-HT090
C 15	330.4	24.8	376	9 AA488304	AA488304 ae30a04.r
C 16	322	24.2	386	14 BQ359015	BQ359015 PM3-HT090
C 17	309.2	23.2	352	14 BQ359019	BQ359019 PM3-HT090
C 18	293.2	22.0	324	12 BF986192	BF986192 RC3-GN004
C 19	280.8	21.1	417	14 BQ359018	BQ359018 PM3-HT090
C 20	280	21.0	333	12 BF832937	BF832937 PM3-HT090
C 21	266.6	20.0	295	10 AW135014	AW135014 UI-H-Bil-
C 22	262.4	19.7	279	12 BF195348	BF195348 7n21a12.x
C 23	252.6	19.0	365	10 BE089961	BE089961 RC6-BT070
C 24	249.8	18.8	517	12 BG188342	BG188342 RST7360 A
C 25	248.8	18.7	470	12 BG200817	BG200817 RST20025
C 26	246.2	18.5	404	9 AA376713	AA376713 EST89177
C 27	245.2	18.4	405	9 AA376719	AA376719 EST89177
C 28	242.4	18.2	519	10 AW951480	AW951480 EST363550
C 29	239	18.0	250	9 AI630950	AI630950 ty99g10.x
C 30	238.4	17.9	410	14 T29650	T29650 EST89132 Hu
C 31	236	17.7	375	9 AA376681	AA376681 EST89134
C 32	228.6	17.2	335	9 AA376714	AA376714 EST89171
C 33	222	16.7	257	12 BF832847	BF832847 PM3-HT090
C 34	211	15.9	457	9 AA376682	AA376682 EST89135
C 35	203.4	15.3	306	10 BE163886	BE163886 QV3-HT046
C 36	197.8	14.9	426	9 AA376680	AA376680 EST89133
C 37	181.4	13.6	302	9 AA376694	AA376694 EST89151
C 38	179.2	13.5	341	9 BQ309679	BQ309679 RC6-BT070
C 39	175.6	13.2	329	14 BQ309679	BQ309679 RST25360
C 40	169.6	12.7	321	12 BG205925	BG205925 RST25360
C 41	163	12.2	258	12 BF834202	BF834202 PM3-HT090
C 42	161.4	12.1	314	10 AW409123	AW409123 sal1_a7 S
C 43	161	12.1	209	12 BF835534	BF835534 PM3-HT090
C 44	155.6	11.7	173	12 BF088945	BF088945 PM3-HT090
C 45	155	11.6	172	12 BF944749	BF944749 PM3-NN117

ALIGNMENTS

RESULT 1
AI224097/c
LOCUS q135a03.x1 Soares_Nhmpu_s1 Homo sapiens cDNA clone IMAGE:1858444
DEFINITION 3', mRNA sequence.
ACCESSION AI224097
VERSION AI224097.1 GI:3806810
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 876 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1..492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1858444"


```

RESULT 3
AW269813/c
LOCUS
DEFINITION
  xv45f02.x1 Soares_NFL_T_GBC_S1 Homo sapiens linear EST 03-JAN-2000
  IMAGE:2816091 3', mRNA sequence.
ACCESSION
  AW269813
VERSION
  AW269813.1 GI:6656843
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (Info@image.llnl.gov) for further information.
  Seq primer: -40UP from Gibco
  High quality sequence stop: 462.
FEATURES
  source
    1..523
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2816091"
    /lab_host="Soares_NFL_T_GBC_S1"
    /lab_host="DH10B"
    /notes="Organ: pooled; Vector: pMT73D-Pac (Pharmacia) with
    a modified polylinker; Site_1: Not I; Site_2: Eco RI;
    Equal amounts of plasmid DNA from three normalized
    libraries (fetal lung MbHL19W, testis NHT, and B-cell
    NCI-CGAP_GCB1) were mixed, and ss circles were made in
    vitro. Following HAP purification, this DNA was used as
    tracer in a subtractive hybridization reaction. The driver
    was PCR-amplified cDNAs from pools of 5,000 clones made
    from the same 3 libraries. The pools consisted of
    I.M.A.G.E. clones 297480-302087, 682632-687239,
    726408-728711, and 729096-731399. Subtraction by Bento
    Soares and M. Fatima Bonaldo. "
    177 a 109 c 85 g 152 t
BASE COUNT
  177 a 109 c 85 g 152 t
ORIGIN

Query Match 35.9%; Score 478.4; DB 10; Length 523;
Best Local Similarity 99.8%; Pred. No. 4.3e-102;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACGTAATTT 60
DB 481 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACGTAATTT 422
QY 61 ACAGTCTTTACAAATGCTCGGAATTTTGACATAGGGAATGTTAAGGTTACTTGGC 120
DB 421 ACAGTCTTTACAAATGCTCGGAATTTTGACATAGGGAATGTTAAGGTTACTTGGC 362
QY 121 TGGAAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTTACGATGAGGGGGAATTTG 180
DB 361 TGGAAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTTACGATGAGGGGGAATTTG 302
QY 181 AGGCCCTTAAGGCTAAACAAATAATCATGATCTGAGATAGTGGCTAATGTCGCTCCCA 240
DB 301 AGGCCCTTAAGGCTAAACAAATAATCATGATCTGAGATAGTGGCTAATGTCGCTCCCA 242
QY 241 GGCCTAAATTTGGACACGTTTTTCTGATCTGCTTTGAGAGTACTTTCTTTGACAGAAA 300
DB 241 GGCCTAAATTTGGACACGTTTTTCTGATCTGCTTTGAGAGTACTTTCTTTGACAGAAA 182
QY 301 TTTTCATCTCTGCTGCTGCTGCTATATCTCCCTTTATAGGAGCCATTTGATTTCTTCC 360
DB 181 TTTTCATCTCTGCTGCTGCTGCTATATCTCCCTTTATAGGAGCCATTTGATTTCTTCC 122

```

```

QY 361 TTTTGTGGGAAATGTCACATTTTCCATTTTTCAGATCTTTTGTGCTGCTACTAATGCCATTAT 420
DB 121 TTTTGTGGGAAATGTCACATTTTCCATTTTTCAGATCTTTTGTGCTGCTACTAATGCCATTAT 62
QY 421 TGGTAATGCCGTTATTTGGTGAATACACATAGTAAATAAATCTGTTACAGTAATCTTACA 480
DB 61 TGGTAATGCCGTTATTTGGTGAATACACATAGTAAATAAATCTGTTACAGTAATCTTACA 2

RESULT 4
BG121534
LOCUS
DEFINITION
  602352872F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4451201 5',
  mRNA sequence.
ACCESSION
  BG121534
VERSION
  BG121534.1 GI:12615043
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM10237 row: p column: 18
  High quality sequence start: 2
  High quality sequence stop: 680.
FEATURES
  source
    1..888
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4451201"
    /lab_host="NIH_MGC_90"
    /tissue_type="adenocarcinoma, cell line"
    /notes="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
    Average insert size 1.7 kb. Library enriched for
    full-length clones and constructed by Life Technologies.
    Note: this is a NIH_MGC Library."
    240 a 161 c 198 g 289 t
BASE COUNT
  240 a 161 c 198 g 289 t
ORIGIN

Query Match 30.4%; Score 405; DB 12; Length 888;
Best Local Similarity 89.8%; Pred. No. 7.9e-85;
Matches 468; Conservative 0; Mismatches 50; Indels 3; Gaps 3;

QY 1 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACGTAATTT 60
DB 284 CTATGTTTTAGAAATCAAGATGAACCGGTAAAGCTGTCATGTCACCAACGTAATTT 343
QY 61 ACAGTCTTTACAAATGCTCGGAATTTTGACATAGGGAATGTTAAGGTTACTTGGC 120
DB 344 ACAGTCTTTACAAATGCTCGGAATTTTGACATAGGGAATGTTAAGGTTACTTGGC 403
QY 121 TGGAAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTTACGATGAGGGGGAATTTG 180
DB 404 TGGAAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTTACGATGAGGGGGAATTTG 463
QY 181 AGGCCCTTAAGGCTAAACAAATAATCATGATCTGAGATAGTGGCTAATGTCGCTCCCA 239
DB 464 AGGCCCTTAAGGCTAAACAAATAATCATGATCTGAGATAGTGGCTAATGTCGCTCCCA 523

```

```

QY 240 AGGCCTAATTTGGGAACAGTTTTCTCGATGCTTTGAGAGTACTTCTTTTGACAGAA 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 AGG-CTAATTTGGGAACAGTTTTCTCGATGCTTTGAGAGTACTTCTTTTGACAGAA 582
QY 300 ATTTCATTCGCTTGGCCATTCATATCTCCCTTTATAGGAGCCATTCGATTTCTTTC 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 583 ATTTCATTCGCTTGGCCATTCATATCTCCCTTTATAGGAGCCATTCGATTTCTTTC 641
QY 360 CTTTGTGGGAATGTCCTCATTCAGATTTTCAGATCTTTTGTGATGTCGACTTAATGCCATT 701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 642 CTTTGTGGGAATGTCCTCATTCAGATTTTCAGATCTTTTGTGATGTCGACTTAATGCCATT 701
QY 420 TTGTAATGCCGTTATTGGTCAATACAGCATAGTTAAATAAAGCTGTACAGTAATCTAC 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 702 ATGTAATGCCGTAATGGGAATCACGCATGTTTATCAACCGTCCCGGGGACATTTAC 761
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 ACTTGGATTTCGTCACCTTACCATAGGCTTTTGAATGA 520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 762 ACTTGGATTTCGTCACCTTACCATAGGCTTTTGAATGA 520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
BF831772/c 427 bp mRNA linear EST 13-JAN-2001
LOCUS PM3-HT0909-151000-009-e12 HT0909 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF831772
ACCESSION BF831772.1 GI:12179873
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 427)
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-HT0909-
151000-009-e12&t3=2000-10-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 427.
Location/Qualifiers
1. 427
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0909"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site: 1; SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
146 a 95 c 72 g 113 t 1 others
BASE COUNT
ORIGIN

```

```

Query Match 30.0%; Score 399; DB 12; Length 427;
Best Local Similarity 99.8%; Pred. No. 2.le-83;
Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 42 TGTACCAACCTGAAATTTACAGTGTTCACAAATGCTGGAAATTTTGCACATGCCATAGG 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 TGTACCAACCTGAAATTTACAGTGTTCACAAATGCTGGAAATTTTGCACATGCCATAGG 368
QY 102 AATGTTAAGTACTTGGCTGGAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTAG 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 AATGTTAAGTACTTGGCTGGAATTTATCAGACTTGTGAGTAAACAAGTTGAAGTTAG 308
QY 162 CAGATGAGGGGGAATATTGAGGCCCTAAGCTAAACAATAATATCAGTATCTGAGATAG 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 CAGATGAGGGGGAATATTGAGGCCCTAAGCTAAACAATAATATCAGTATCTGAGATAG 248
QY 222 TGGCTAATGTGGCTCCCGCCAGGCTTAATTTGGGAACAGTTTTTCTGATTCCTTTGAGAAG 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 TGGCTAATGTGGCTCCCGCCAGGCTTAATTTGGGAACAGTTTTTCTGATTCCTTTGAGAAG 188
QY 282 TACTTTCTTTTGACAGAAATTTTCATCTGCTTGGCAATTCCTATATTCCTCTTATAGG 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 TACTTTCTTTTGACAGAAATTTTCATCTGCTTGGCAATTCCTATATTCCTCTTATAGG 128
QY 342 AGCCATTGGATTTCTTTCTCTTTTGTGGGAATGTCCTATAGCATTTTTCAGATCTTTTGA 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 AGCCATTGGATTTCTTTCTCTTTTGTGGGAATGTCCTATAGCATTTTTCAGATCTTTTGA 68
QY 402 TGTGCACTAATGCCATTTTGGTAATGCCGTTAATGGTGA 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 TGTGCACTAATGCCATTTTGGTAATGCCGTTAATGGTGA 28

RESULT 6
BF677736/c 417 bp mRNA linear EST 20-SEP-2000
LOCUS RC3-GN0042-100800-011-e04 GN0042 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF677736
ACCESSION BF677736.1 GI:10221394
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 417)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC3-GN0042-100
800-011-e04&t3=2000-08-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 37
High quality sequence stop: 417.
Location/Qualifiers
1. 417

FEATURES
source

```


Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=PM3&t2=PM3-HT0909-181000-010-g05&t3=2000-10-18&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 39

High quality sequence stop: 403.

FEATURES

source

1. .403

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0909"

/dev_stage="Adult"

/note="Organ: head neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 105 a 68 c 88 g 142 t

ORIGIN

Query Match 27.8%; Score 369.4; DB 12; Length 403;

Best Local Similarity 99.7%; Pred. No. 1.9e-76;

Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 68 TTACAAATGCTCGGAATTTGCACTGCCATAGGGAATGTTAAGTTACTTGGCTGGAAAT 127

Db 33 TTACAAATGCTCGGAATTTGCACTGCCATAGGGAATGTTAAGTTACTTGGCTGGAAAT 92

QY 128 TATCAGACTGTGAGTAACAGTGAAGTTAGCAGATGAGGGGAATATTGAGGCCCC 187

Db 93 TATCAGACTGTGAGTAACAGTGAAGTTAGCAGATGAGGGGAATATTGAGGCCCC 152

QY 188 TAAGCTAAACAAATAATCATGATCTGAGATAGTGGCTATGCTGCCCGCCGCTAA 247

Db 153 TAAGCTAAACAAATAATCATGATCTGAGATAGTGGCTATGCTGCCCGCCGCTAA 212

QY 248 TTGGGAACAGTTTTCCTGATTCCTTTGAGAACTACTTCTTTTGACAGAAATTTTCAT 307

Db 213 TTGGGAACAGTTTTCCTGATTCCTTTGAGAACTACTTCTTTTGACAGAAATTTTCAT 272

QY 308 TCTGCTGCCATTGCTATATTCCTCCCTTTATAGAGCCATTGGATTTCTTCTTTTGTG 367

Db 273 TCTGCTGCCATTGCTATATTCCTCCCTTTATAGAGCCATTGGATTTCTTCTTTTGTG 332

QY 368 GGAATCTCCCATTAGCATTTTCAGATCTTTTGTATGTCACATAATGCCATTATGTAAT 427

Db 333 GGAATCTCCCATTAGCATTTTCAGATCTTTTGTATGTCACATAATGCCATTATGTAAT 427

QY 428 GCCGTTATTGG 438

Db 393 GCCGTTATTGG 403

RESULT 9

LOCUS BG011354/c

DEFINITION RCI-GN0268-091200-011-e10 GN0268 Homo sapiens CDNA, mRNA sequence.

ACCESSION BG011354

VERSION BG011354.1 GI:12459469

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC1&t2=RC1-GN0268-091200-011-e10&t3=2000-12-09&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 10

High quality sequence stop: 408.

FEATURES

source

1. .408

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="GN0268"

/dev_stage="Adult"

/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 142 a 90 c 71 g 105 t

ORIGIN

Query Match 26.5%; Score 353; DB 12; Length 408;

Best Local Similarity 96.8%; Pred. No. 1.4e-72;

Matches 392; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

QY 42 TGTACCAACGTGAAATTTACAGTGTGTTTACAAATGCTTGGAAATTTTGCACTGCCATAGGG 101

Db 408 TGTACCAACGTGAAATTTACAGTGTGTTTACCAATGCTTGGAAATTTTGCACTGCCATAGGG 349

QY 102 AATGTTAAGTTACTTGGCTGGAATTTATCAGACTTGTGAGTAACAAGTTGAAGTTTATG 161

Db 348 AATGTTAAGTTACTTGGCTGGAATTTATCAGACTTGTGAGTAACAAGTTGAAGTTTATG 289

QY 162 CAGATGAGGGGAATATTGAGGCCCTTAAGCTAAACAAATATATCATGATCTGAGATAG 221

Db 288 CAGATGAGGGGAATATTGAGGCCCTTAAGCTAAACAAATATATCATGATCTGAGATAG 229

QY 222 TGGTAATATGCGTCCCGAGCCCTTAATTTGGGAACAGCTTTTCTGCTGATTCCTTTGAGAAG 281

Db 228 TGGTAATATGCGTCCCGAGCCCTTAATTTGGGAACAGCTTTTCTGCTGATTCCTTTGAGAAG 169

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10228 row: n column: 16
High quality sequence stop: 675.
Location/Qualifiers
1. .951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:447695"
/lab_host="NIH_MGC_89"
/tissue_type="hypermethylation, cell line"
/note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.3 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
281 a 221 c 196 g 253 t

FEATURES
source

BASE COUNT 281 a 221 c 196 g 253 t
ORIGIN
Query Match 25.2%; Score 335.2; DB 12; Length 951;
Best Local Similarity 95.8%; Pred. No. 2.1e-68;
Matches 366; Conservative 0; Mismatches 13; Indels 3; Gaps 2;
QY 709 GACAGAGACTCTGCTCAAAAAGGACATTTATCATATTAACATCTTATTAGACCC 768
Db 76 GACAGAGACTCTGCTCAAAAAGGACATTTATCATATTAACATCTTATTAGACCC 135
QY 769 TAATTTCTTATCTGAAGGACATCTTTTAAACAGTTAAGTACGATGCTCAACAG 828
Db 136 TAATTTCTTATCTGAAGGACATCTTTTAAACAGTTAAGTACGATGCTCAACAG 193
QY 829 ACAATATTTCTGATCAGATAGTCCCTGCTCAACAGTAGCAATGTGTTTCATAAAGTG 888
Db 194 ACAATATTTCTGATCAGATAGTCCCTGCTCAACAGTAGCAATGTGTTTCATAAAGTG 253
QY 889 GGAAGAAACACAGATTTTAAAGTAACTTTTGGGACAGTATGAGTAATAATAAACT 948
Db 254 GGAAGAAACACAGATTTTAAAGTAACTTTTGGGACAGTATGAGTAATAATAAACT 312
QY 949 CTGGTCTCCCTTAAGAAAAAACCCTCCACCTTTACTGTGTCATTTATATCCCTTA 1008
Db 313 CTGGTCTCCCTTAAGAAAAAACCCTCCACCTTTACTGTGTCATTTATATCCCTTA 372
QY 1009 GTTCAAAGTTAATTAATCTTATCTGATATGCTTTTATACCAAGAGCCTTATCAGC 1068
Db 373 GTTCAAAGTTAATTAATCTTATCTGATATGCTTTTATACCAAGAGCCTTATCAGC 432
QY 1069 CAGTTCCAGAACCAACCTATA 1090
Db 433 CTTGTACTACAGTATCTTTA 454

RESULT 14
BF832034
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 396)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunschein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed

RESULT 15
AA488304
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 396)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunschein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM3&t2=PM3-HT0909-181000-011-h06&t3=2000-10-18&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 87.
Location/Qualifiers
1. .396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0909"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
104 a 57 c 96 g 139 t

FEATURES
source

Query Match 25.0%; Score 333; DB 12; Length 396;
Best Local Similarity 96.2%; Pred. No. 7e-68;
Matches 384; Conservative 0; Mismatches 10; Indels 5; Gaps 4;
QY 42 TGTACCAACGTTAAATTTACAGTCTTTACAAATCTCTGGAATTTTGGCACTGCCATAGG 101
Db 1 TGTACCAACGTTAAATTTACAGTCTTTACAAATCTCTGGAATTTTGGCACTGCCATAGG 60
QY 102 AATGTTAAGTTACTTGGCTGGAATTTATCAGACT--TGTGAGTAAACAAGTTGAAGTTT 159
Db 61 AATGTTAAGG-TACTTGGGTGGAATTTATTAGAAATTTGCTGAGTAAACAAGTTGAAGTTT 119
QY 160 AGCAGATCAGGGGGAATTTAGAGGCCCTTAAGGCTTAACAAATAATCAGTATCTCAGAT 219
Db 120 AGCAGATCAGGGGGAATTTAGAGGCCCTTAAGGCTTAACAAATAATCAGTATCTCAGAT 178
QY 220 AGTGCTAATGTGCTCCCGAGGCTAATTTGGGAACAGTTTTCCTGATTTGAGA 279
Db 179 AGTGCTAATGTGCTCCCGAGGCTAATTTGGGAACAGTTTTCCTGATTTGAGA 238
QY 280 AGTACTTTCTTTTACAGAAATTTTCATTCGTTGCCATTCGCTATATTCCTCTTTATA 339
Db 239 AGTACTTTCTTTTACAGAAATTTTCATTCG-TGTGGAATTCGCTATATTCCTCTTTATA 297
QY 340 GGAGCCATTGGATTTCTTTCTCTTTTGTGGGAAATGTCCTTATAGCATTTTCAGATCTTTT 399
Db 298 GGAGCCATTGGATTTCTTTCTCTTTTGTGGGAAATGTCCTTATAGCATTTTCAGATCTTTT 357
QY 400 GATGTGCACTAATGCCATTTATTTGGTAATGCCGTTATTGG 438
Db 358 GATGTGCACTAATGCCATTTATTTGTAATGCGGTTATTGG 396

RESULT 15
AA488304
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 396)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunschein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed

```

BASE COUNT      97 a      77 c      75 g      127 t
ORIGIN

Query Match      24.8%; Score 330.4; DB 9; Length 376;
Best Local Similarity 95.5%; Pred. No. 2.9e-67;
Matches 340; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 370 AAATGTCCCATTAGCATTTTCAGATCTTTTGATGTGCACATAATGCCATTATTCGTTAATGC 429
Db 1 AAATGTCCCATTAGCATTTTCAGATCTTTTGATGTGCACATAATGCCATTATTCGTTAATGC 60
QY 430 CGTTATTTGGTGAATACAGCATAGTTAAATAAAGCTTTACAGTAAATCTACACTTGGATTT 489
Db 61 CGTTATTTGGTGAATACAGCATAGTTAAATAAAGCTTTACAGTAAATCTACACTTGGATTT 120
QY 490 GTGTCACCTCTACCAATAGCCTTTTCAATGACTGAAAGTGTAAACAGAGAAAGAGGCATG 549
Db 121 GCTGACACCTCTACCAATAGCCTTTTGAATGACTGAAAGTGTAAACAGAGAAAGAGGCATG 180
QY 550 TCTGCAGAAAGAGATAGCTAAATATTTTGGTACTTTATCTGAAATCCAAAGATGCTGCTT 609
Db 181 TCTGCAGAAAGAGATAGCTAAATATTTTGGTACTTTATCTGAAATCCAAAGATGCTGCTT 240
QY 610 CCCCTGCAGGTGTTTCTCTCTTACGATCCATTTGAATCCCTCTGGGAGCAGAGAC 669
Db 241 CCCCTGCAGGTGTTTCTCTCTTACGATCCATTTGAATCCCTCTGGGAGCAGAGAC 300
QY 670 AGTTAGTAGAACTCTCCATTCTCTTGTTTGGTTTTTAAGACAGAGACTCTGCTTC 725
Db 301 AGTTAGTAGAACTCTCCATTCTCTTGTTTGGTTTTTAAGACAGAGACTCTCTGCTTC 356

```